MECHANISMS OF TELOMERASE RNA INSUFFICIENCY IN DISEASE

by

Dustin L. Gable

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
April 2019

Abstract

Incremental decreases in the low abundance telomerase RNA (TR) cause short telomere syndromes, the most common of premature aging disorders. Thirteen genes have been shown to cause the short telomere syndrome phenotype (STS) and six of them (46%) disturb telomerase RNA (TR). In addition to *TR* itself, mutations in *DKC1*, *NOP10*, *NHP2*, *PARN*, and as we recently found, *NAF1,* all disturb TR integrity pointing to TR abnormalities as the most commonly perturbed pathway in human short telomere syndromes. To test how mutations in *NAF1,* a box H/ACA RNA ribonucleoprotein, cause phenotypes such as pulmonary fibrosis, emphysema and bone marrow failure, we used CRISPR/Cas9 to disrupt the *Naf1* locus in mice and found that *Naf1* was essential. However, *Naf1*$^{+/-}$ mice survived and showed haploinsufficiency for TR. Moreover, *Naf1*$^{+/-}$ mice showed a similar decrease of sno- and sca- H/ACA RNA levels as with TR. However, we did not observe any phenotypic abnormalities in these mice including no phenotypic evidence of a ribosomopathy such as hematologic, skeletal or neurologic defects. In a high resolution analysis of 82 ribosomal pseudouridylation sites in ribosomal RNA components, 28S, 18S, 5.8S, using Pseudo-Seq, we found no abnormalities in *Naf1*$^{+/-}$ mice. These data indicate that although the levels of box H/ACA RNAs are compromised in *Naf1*$^{+/-}$ mice, they are sufficient to direct pseudouridylation and maintain ribosome function. In contrast, NAF1 mutations cause TR haploinsufficiency, telomere shortening, and telomere-mediated disease. To identify additional novel genetic causes of STS and TR insufficiency we used genome-wide linkage and identified heterozygous loss-of-function mutation in ZCCHC8, a nuclear RNA exosome targeting component. We show that ZCCHC8 is required for telomerase function by targeting a longer nascent TR to the RNA exosome for 3'end maturation. *Zcchc8*$^{+/-}$ mice had low TR levels and defective TR 3'ends, but their transcriptome was otherwise preserved. In contrast, *Zcchc8*$^{-/-}$ mice showed a progressive fatal neurodevelopmental phenotype that resembled a ciliopathy. Their transcriptome was highly dysregulated with 3'end misprocessing of low abundance intronless RNAs that were similar in length to TR including replication-dependence histones. Our data underscore the dosage sensitivity for telomere maintenance not only to TR levels but also to the levels of some of its biogenesis factors and indicate that vertebrate TR shares a 3'end processing pathway with a subset of low abundance RNAs, including replication-dependent histones.

Primary reader: Mary Armanios

Secondary Reader: Carol W. Greider, Ph.D.

Acknowledgments

The ideas and discoveries presented hereto reflect the insight from a team of past and present members of the Armanios lab and collaborations across institutions. Additionally, this work would not have been possible without the continued support and thoughtful discussion with the laboratory of Carol Greider.

I'd like to thank the Johns Hopkins MSTP program and Human Genetics Program for the educational foundation, which guided this work and the Johns Hopkins University Core Facilities for providing invaluable resources.

I would like to thank my friends and family for their steadfast encouragement and unwavering support of my research endeavors. To my Aunt Barbra Deepe whose resilience in her battle with cancer and appreciation for life continue to motivate me along this journey. To my late Aunt Mildred Bidlack for instilling in me as a child a reverence for education. And finally to my grandmother Dolores Lammers for her unconditional love and confidence in me to persevere and enjoy each step of the way.

Finally, I would like to thank my mentor Mary Armanios who has been pivotal in my development during this time, for setting an example as a physician scientist while instilling in me virtue in science and society that guide our decisions, and for her belief in our work for our community and families as critical to the successful intersection of science and medicine that has shaped me as a person, clinician, and scientist.

Table of Contents

ABSTRACT .... **II**

ACKNOWLEDGMENTS .... **IV**

TABLE OF CONTENTS .... **V**

LIST OF TABLES .... **VII**

LIST OF FIGURES .... **VIII**

CHAPTER ONE INTRODUCTION **1**

CHAPTER TWO HAPLOINSUFFICIENCY OF THE H/ACA BIOGENESIS FACTOR *NAF1* SELECTIVELY DISRUPTS TELOMERE LENGTH MAINTENANCE WHILE SPARING RIBOSOMAL RNA FUNCTIONS **11**

- INTRODUCTION 12
- RESULTS 13
- DISCUSSION 26
- METHODS 29

CHAPTER THREE ZCCHC8 IS MUTATED IN PULMONARY FIBROSIS **36**

- INTRODUCTION 37
- RESULTS 39
- DISCUSSION 49
- METHODS 51

CHAPTER FOUR LOSS OF ZCCHC8 IN MICE LEADS TO SEVERE NEURODEVELOPMENTAL DEFECTS **60**

- INTRODUCTION 61
- RESULTS 62
- DISCUSSION 74
- METHODS 77

REFERENCES **84**

CURRICULUM VITAE **89**

EDUCATION **89**

RESEARCH **89**

PUBLICATIONS 89

PRESENTATIONS 90

AWARDS 92

List of Tables

Table 1. Dose-dependent manifestations of mutations in TR and TR proteins………......4

Table 2. TR processing characteristics in cilia, yeast, and humans.................................5

List of Figures

Figure 1. Genes identified as regulators of telomere length in yeast. Unique genes were identified as having roles in telomere length regulation after review among three independent gene deletion, haploid, and essential gene screens for telomere length maintenance in *Saccharomyces cerevisiae (12-14).* A total of 166 genes were grouped into 10 categories and the frequency of genes in each category is represented here in this pie chart. 2

Figure 2. Model for hTR biogenesis and processing. hTR is transcribed from its own promoter (right angle arrow) by RNA polymerase II (pol II). Post-transcriptional stability is provided by binding of the Dyskerin complex to the TR box H/ACA motif. The Dyskerin complex also includes NAF1, NHP2, and NOP10. End processing at the 5' and 3' end is shown (large curved arrows). TGS1 trimethylates the 5' trimehthylguanosine cap (TMG) which may be coupled to the cap binding complex (CBC) to assist in 3' end processing (dashed arrow). Extended nascent TR transcripts as long as or greater than 1000 nucleotides (nt) are either targeted for maturation via binding by ZCCHC8 and the NEXT complex to the RNA exosome bound by the exonuclease DIS3 in the nucleoplasm or targeted for degradation via the TRAMP complex to the RNA exosome in the nucleolus. After trimming by the RNA exosome, shorter adenylated nascent TR forms (abbreviated with line and short polA tail) are processed further by the deadenylase PARN to the most common mature TR form that is 451 nt and bound by the Dyskerin complex. NAF1 traffics TR to the Cajal body where it is exchanged with GAR1. 10

Figure 3. TR levels are decreased in mouse embryonic fibroblasts with decreased NAF1 levels. a. Schema of mouse wild-type *Naf1* locus and CRISPR/Cas9 editing strategy to knockout *Naf1.* A sgRNA was designed to target exon 1 and introduce a nonsense mutation via homology directed recombination that would also alter the native *SacII* restriction site to a correctly edited *NaeII* restriction site that. Restriction digest of the target region PCR product with either enzyme followed by gel electrophoresis could be used to detect the presence of a wildtype or correctly edited allele. **b.** Immunoblot of lysate from human embryonic kidney (HEK) 293FT cells transfected with Myc-tagged Naf1 cDNA-containing plasmid at shown concentrations. Immunoblot for human NAF1 (hNAF1) and mouse NAF1 and Myc shows specificity of the murine NAF1 antibody. **c.** Immunoblot for NAF1 in CRISPR edited mouse embryonic fibroblasts. **d.** TR measured by qRT-PCR in *Naf1*$^{+/+/+}$ (blue) MEFs compared to two different *Naf1*$^{+/-/-}$ mutant MEF cell lines (pink and purple) each containing one allele with an in frame mutation and two with framseshift mutations. Levels were relative to hypoxanthine-guanine phosphoribosyltransferase (Hprt). For both *Naf1*$^{+/+/+}$ versus *Naf1*$^{+/-/-}$ comparisons, $P < 0.001$, n=3 technical replicates. 15

Figure 4. *Naf1*$^{+/-}$ mice were created using CRISPR/Cas9 editing. a. Genotyping using the *SacII* and *NaeI* restriction digest strategy (Figure 3a) of C57BL/6J mice created from injection of Cas9 protein, sgRNA targeting mouse *Naf1* exon 1, and an oligo for homology directed recombination into single cell zygotes compared to control wildtype MEFs (n=6 mice). b. Schema of wild-type *Naf1* locus and CRISPR/Cas9-induced 322-bp deletion (Δ322) in exon 1 and the amino acid sequence created by the frameshift mutation created (blue=wildtype, pink=mutant). c. Immunoblot for NAF1 in adult skin fibroblasts relative to Actin. 17

Figure 5. RNA Levels in Naf1$^{+/-}$ mice. a. TR measured by qRT-PCR in Naf1$^{+/+}$, Naf1$^{+/-}$, mTR$^{+/-}$, and mTR$^{-/-}$ mice (total spleen RNA, n = 3 mice per group) and normalized to hypoxanthine-guanine phosphoribosyltransferase (Hprt). b-c. Northern blot for mTR (mouse TR) normalized to 5.8S (b) with quantifications of Naf1$^{+/+}$(blue), Naf1$^{+/-}$ (pink), mTR$^{+/-}$ (purple), and mTR$^{-/-}$ below (c). d. H/ACA, C/D, splicing (snRNA), long noncoding (lnRNA) and messenger (mRNA) RNA levels measured by qRT-PCR (n = 3 to 8 mice per group, total spleen RNA) and normalized to Hprt in Naf1$^{+/+}$(blue), Naf1$^{+/-}$ (pink). Error bars represent SEM. $^{**}P < 0.01$ and $^{***}P < 0.001$ (Student's t test). 19

Figure 6. Telomere length in Naf1$^{+/-}$ mice. a. Generational breeding strategy for generating *Naf1*$^{+/-}$G1, *Naf1*$^{+/-}$G2, *Naf1*$^{+/-}$G3, and *Naf1*$^{+/-}$G4 mice for telomere length measurement. **b.** Relative telomere

length in lymphocytes isolated from spleens of *Naf1*$^{+/+}$ G1 (blue), *Naf1*$^{+/-}$ G1 (pink), *Naf1*$^{+/-}$ G2 (purple), *Naf1*$^{+/-}$ G3 (yellow), *Naf1*$^{+/-}$ G3 (green), and *mTR*$^{-/-}$G4 (orange) mice measured by measured by flow cytometry and fluorescence in situ hybridization (flow-FISH) (n=3 mice per group). **c.** Relative weight of testes and spleens isolated from mice whose telomere lengths were measured in *b* (n=3 mice/ group). 21

Figure 7. Pseudouridylation is intact in *Naf1*$^{+/-}$ mice. a. Alignment and Pseudo-seq strategy for identification of 82 pseudouridylation sites in mouse rRNA. b. Scatter plot of average Pseudo-seq signal ratio in *Naf1*$^{+/+}$ (n=4) versus *Naf1*$^{+/-}$ (n=3) mice graphed for 82 sites in three rRNA subunits. 23

Figure 8. Schema of mouse box H/ACA snoRNAs quantified showing the complementary guide sequence and the rRNA pseudouridine (Ψ) target site (red). a-i. The six snoRNA quantified in Fig 2.3d are shown here with their complementary sequence targeting 9 rRNA sites (shown in panels a-i). Each rRNA respective subunit is shown below. Eight of these rRNA sites were detected by Pseudo-seq as annotated in Table 1. In panels *d, g,* and *i* (designated by *), the murine RNA reference sequence was incomplete at the 5' end, and sequence complementarity was derived from the genomic sequence based on build GRCm38/mm1. 24

Figure 9. Linkage analysis identifies the *ZCCHC8* locus containing a novel mutation that segregates with familial pulmonary fibrosis and telomerase RNA (*TR*) insufficiency. A. Pedigree with proband (designated by arrow) and relatives who had pulmonary fibrosis are indicated by the shaded symbols (Key). The '?' refers to individuals whose clinical status was uncertain at the time of clinical assessment and the individuals shaded in gray had an unknown cause of death. *Refers to individuals for whom DNA was available. B. *TR* levels measured by quantitative real time PCR (qRT-PCR) in lymphoblastoid cell lines (LCLs). Arrow refers to proband (red dot) and pedigree identifiers refer to A (proband's sons). *TR* levels from a *DKC1* mutation carrier is a positive control. The data represent a mean of three experiments, each from independent RNA isolations. C. Telogram shows age-adjusted lymphocyte telomere length by flow cytometry and fluorescence in situ hybridization (flowFISH) in the proband (arrow) and his family. Pedigree designations are as in A. The telogram is based on clinically validated dataset from 192 controls. D. Phenotype assignments as shown in key that were used in genome wide linkage analysis based on the presence of pulmonary fibrosis as well as *TR* levels (B) and telomere length measurement (C). The *ZCCHC8* single nucleotide variant for the P186 change is also shown below each pedigree symbol and italicized genotypes refer to obligate carriers. E. Log of the odds ratio score across the autosomal chromosomes calculated from 15 individuals identified by * in (A). The linkage peak on chromosome 12 is denoted by the arrow. F. Novel protein coding mutation identified p.P186L falls in an unstructured domain of *ZCCHC8* with alignment across eight vertebrate species shown. The darker shading denotes more conserved identity conserved residue identity. CCHC refers to Zinc-knuckle domain; PSP refers to proline-rich domain. 40

Figure 10. Pulmonary fibrosis-associated mutation affects ZCCHC8 stability. A. Immunoblot of endogenous ZCCHC8 levels in lymphoblastoid cell lines (LCLs) from healthy controls (C1 and C2), unaffected relatives and mutation carriers as indicated. Result replicated in two other blots. Pedigree identifiers refer to Figure 1A. B. Immunoblot showing ZCCHC8, SKIV2L2 and RBM7 levels in primary skin fibroblasts from the proband. C. Total *ZCCHC8* mRNA levels in LCLs from unaffected relatives (n=4) and *ZCCHC8* p.P186L mutation carriers (n=3). D. Chromatogram showing *ZCCHC8* p.P186L mutation is detectable in mRNA from LCLs. E. Immunoblot of transfected Myc-tagged (into 293FT cells) and endogenous ZCCHC8 showing the P186L mutation is less stable. F. Immunoblot showing efficacy of shRNA knockdown of Luciferase (Luc), ZCCHC8 and NAF1 in HeLa cells on protein levels. G. Total *TR* levels measured by qRT-PCR after knockdown. Mean is from 3 independent knockdowns. H. Northern blot shows low *TR* levels after stable knockdown of *ZCCHC8* and *NAF1*. Quantification is shown below each of the gels and each experiment was replicated at least twice where not noted. For C and G, the mean ± SEM is shown. **$P<0.01$ and ***$P<0.001$ (Student's *t*-test, two-sided)..43

Figure 11. Pulmonary fibrosis-associated mutation affects ZCCHC8 stability. A. Chromatogram showing compound heterozygous frameshift (fs) mutations generated using CRISPR/Cas9 in HCT116 pseudodiploid cell lines. B. Immunoblot for ZCCHC8 in HCT116-edited cells. C. Schema summarizing *TR* 3' Rapid Amplification of cDNA Ends sequencing (3'RACE-seq) method. *TR* 3'ends were generally divided into mature (451 bp) and extended (>451 bp) where extensions are

denoted by gray N's. D. Summary of *TR* 3'RACE-seq fractions in *ZCCHC8*$^{+/+}$ and *ZCCHC8*$^{-/-}$ HCT116 cells. The color-coded key shows the different categories of *TR* forms including adenylated and genomically-extended forms that are adenylated and long extended forms that are >465 nucleotides. Data reflect mean of 3 independent 3'RACE-seq analyses each from a different RNA isolation. *ZCCHC8*$^{+/+}$ cells were isogenic having undergone selection. E. qRT-PCR of extended *TR* forms beyond the 451 mature end (>20, >51, >784 nucleotides). Data reflect mean values from three independent RNA isolations. F. Northern blot for *TR* of edited *ZCCHC8*$^{+/+}$ and *ZCCHC8*$^{-/-}$ HCT116 cells. G. Quantification of six northern blots from three independent RNA isolations. H. Telomerase activity measured by Telomere Repeat Amplification Protocol (TRAP) assay in *ZCCHC8*$^{+/+}$, *ZCCHC8*$^{-/-}$, and *NAF1*$^{S329/S329}$ HCT116 cell extracts. Activity was quantified in serially diluted extracts (1, 1/5, 1/25, 1/125) against a PCR-amplified internal control (IC). An RNAse treated *wild-type* extract and no template PCR reaction are negative controls. I. Mean TRAP activity of 1/5x diluted extracts from 3 independent TRAP assays each from a different cell lysate. J. Summary of 3'RACE-seq of *TR* forms from control and proband primary skin fibroblasts as in (D). K. Mean qRT-PCR values of extended *TR* forms in primary skin fibroblasts as in (E) (n=3 technical replicates). Data are expressed as means ±SEM. *P < 0.05 and **P < 0.01 (Student's *t*- test, two-sided). 46

Figure 12. ZCCHC8 immunoprecipitates *TR* precursors with genomically encoded 3'ends. A. Immunoblot showing input (I) and immunoprecipitated (IP) Myc-tagged ZCCHC8 plasmid transfected at two concentrations into 293FT and harvested at 48h. The lower most panel is an over-exposed blot to show the relative levels of tagged ZCCHC8 relative to endogenous. B. Amplified *TR* from I and IP fractions using primers within the mature *TR* sequence. C. Western blot of IP Myc-tagged ZCCHC8 shows no detectable dyskerin or NAF1 co-IP. D. Deconvoluted *TR* 3'RACE-seq data showing enrichment of genomically encoded *TR* precursors in the Myc-ZCCHC8 fraction compared to input. The genomic TR locus with the 451 mature 3'-end is annotated in the left upper corner. In the color-coded key of the *TR* species, Misc refers to non-genomic, non-templated nucleotides other than 'A.' E. Western blot showing transfection of Myc-tagged *ZCCHC8* and *DIS3* plasmids into CRISPR/Cas9 edited HCT116 cells (Figure 3A). E. qRT-PCR of nascent *TR* that is >51 nucleotides extended beyond the 3' mature *TR* end with transfected tagged ZCCHC8 and DIS3. 48

Figure 13. Design of *Zcchc8* targeting strategy using CRISPR/Cas9. A. Scheme depicts CRISPR targeting of exon 2 within the sgRNA binding site that is 3 base pairs upstream of the PAM sequence to incorporate a 33 base pair insert (arrows) containing a stop codon (red). *BamH1* restriction site and the FLAG sequence identify edited alleles. The screening strategy uses a forward primer in the edited exon and a downstream intronic reverse primer (black arrows). B and C. Immunoblot for ZCCHC8 and SKIV2L2 and RBM7 protein in mouse skin fibroblasts. D and E. Northern blot for mouse *TR* (*mTR*) levels quantified relative to 5.8S rRNA in skin fibroblasts. For (E), mean reflects mice *Zcchc8*$^{+/+}$ (n=4, 2M/2F), *Zcchc8*$^{+/-}$ (n=4, 2M/2F), *Zcchc8*$^{-/-}$ (n=3M), *mTR*$^{+/-}$ (n=2, sex unknown) and *mTR*$^{-/-}$ (n=2, sex unknown). F. *mTR* 3'-extended levels (>20 bp) relative to Hprt as measured by qRT-PCR. Mouse numbers and sex designations are as in E. Data in B-F are from mouse adult ear fibroblasts. G. Genotype frequency from *Zcchc8*$^{+/-}$ interbreeding of E12.5 embryos, newborn pups (P0), and day 30 adult mice with the number of each genotype screened shown. Data are expressed as means ±SEM. *P < 0.05, **P < 0.01, and ***P < 0.0001 (Student's *t*- test, two-sided)................65

Figure 14. ZCCHC8 loss causes a neurogenesis and hydrocephalus defect in adult knockout mice, but heterozygous mice are intact. A. *(Top row)* Images showing head profile of *Zcchc8* wild-type, heterozygous and homozygous null mice (41-46 days old). The labels show the genotype with a male (left) and a female (right) shown for each genotype. *Zcchc8*$^{-/-}$ mice have abnormal head profile with dome cranium and pointed nose as outlined by the dashed line. *(Middle row)* CT head mid-sagittal images show intact skull shape in *Zcchc8*$^{+/-}$ mice, while *Zcchc8*$^{-/-}$ mice dome-shaped crania. *(Bottom row)* Volume-rendered (VR) CT images of mouse calvaria show widened cranial sutures in *Zcchc8*$^{-/-}$ mice (3 of 6 imaged). None of the *Zcchc8*$^{+/+}$ (0 of 4) or *Zcchc8*$^{+/-}$ mice (0 of 4) imaged had this feature. Each vertical group of images is from the same mouse except the last column (2 different females). B. Representative H&E coronal sections from 8-week old male heads show no defects in *Zcchc8*$^{+/-}$ (11 examined) compared to *Zcchc8*$^{+/+}$ mice (10 examined). In contrast, *Zcchc8*$^{-/-}$ mice have severe ventricular dilation (seen in 9 of 11 examined). C. E12.5 brain sections show *Zcchc8*$^{-/-}$ have microcephaly but intact brain structures and no ventriculomegaly. D. E12.5 embryos from a single dam show expected Mendelian ratios but *Zcchc8*$^{-/-}$ embryos have small crania. E. Cranial area of P0

pups measured on VR CT images using ImageJ (*Zcch8*$^{+/+}$ n=7, 3M/4F; *Zcch8*$^{+/-}$ n=10, 3M/7F; *Zcch8*$^{-/-}$ n=4, 3M/1F). F. Cranial area as in E with best fit curve (second order polynomial quadratic) showing *Zcchc8*$^{-/-}$ mice become macrocephalic between birth and adulthood reflecting their hydrocephalus (*Zcchc8*$^{+/+}$, n=4, 3M/1F; *Zcchc8*$^{+/-}$, n=4, 3M/1F; *Zcchc8*$^{+/-}$, n=5, 3M/2F). Dashed lines denote the 95% confidence intervals. F. Total body weight of adult male mice (42 to 63 days old, *Zcch8*$^{+/+}$, n=20; *Zcch8*$^{+/-}$, n=13, *Zcch8*$^{-/-}$, n=5). E and F show mean values ±SEM. **P<0.01 (Student's *t*-test, two-sided). 67

Figure 15. *Zcchc8*$^{-/-}$ mice show an upregulation of low abundance adenylated RNAs including *TR*. A. Heatmap and dendrogram of gene expression showing unsupervised analysis of 9,788 genes. Colors denote mean-subtracted FPKM expression values on a $\log_2$ scale. *Zcchc8*$^{+/+}$ n=5, *Zcchc8*$^{+/-}$ n=3, *Zcchc8*$^{-/-}$, n=6 embryonic brains. Each column is labeled below by WT, HET, KO followed by the embryo number 1,2,3…etc. and refer to respective *Zcchc8* genotypes. The $\log_2$ expression value was subtracted from the mean $\log_2$ expression value of the entire cohort. The dendrogram showing relatedness of the samples is above, and relatedness of the gene transcripts is to the left. The differential change in gene expression is shown as positive and negative change on color scale indicated in key. B and C. Volcano plots depicting the $\log_2$-fold changes (x-axis) vs. $-\log_{10}$ P-values calculated by two-tailed one-way ANOVA of RNA-seq expression data (y-axis) for the *Zcchc8*$^{+/-}$ and *Zcchc8*$^{-/-}$ *vs. Zcchc8*$^{+/+}$ comparisons, respectively. Each dot represents a single transcript. D. Extended reads defined as ≥15 bases beyond the mature RNA end for each of *TR* and *Rmrp* is shown. Data were normalized to total Arf3 reads per embryo and each dot represents values from an embryo. Error bars represent s.e.m and P-value is calculated based on a two-sided Student's t-test. E. Histogram of number of genes at each expression value denoted on the x-axis by the mean $\log_2$FPKM values obtained from *Zcchc8* wild-type embryos (n=5). Genes that have > 2 SD higher levels in the *Zcchc8*$^{+/+}$ vs. *Zcchc8*$^{-/-}$ comparison are shown in red (n=197) and fall on the low abundance end of the histogram with TR and its mean FPKM in wild-type embryos shown. The downregulated genes defined as <2 SD (n=42) are shown in blue and appear uniformly distributed on the distribution......70

Figure 16. Telomerase RNA shares a requirement with replication-dependent histones (RDH) and some cilia genes for Zcch8-dependent 3'-end processing. A. Histogram of the most upregulated (>2SD) transcripts in the *Zcchc8*$^{-/-}$ vs. *Zcchc8*$^{+/+}$ by exon number shows the largest category are intronless (42 of 188 with known gene structure, 22%). The pie charts divides the intronless genes by functional category. **B.** Annotation of 28 upregulated intronless genes (*TR*, histones and cilia) shows a majority of the histones represented are RDH (23 of 24, 96%). The vast majority also have an annotated transcript size in the range of *TR* (NCBI Gene database, https://www.ncbi.nlm.nih.gov/gene, accessed 1/1/2019) between 400 and 560 (22 of 28). The 3' and 5'end analyses denote longer ends at each of these mature RNA boundaries in *Zcchc8*$^{-/-}$ as manually visualized in Integrative Genomic Viewer (IGV). Columns refer to 5'end and 3'end refer to presence of additional reads beyond these boundaries with 5'end reads usually referring to upstream reads that are not necessarily contiguous. **C** and **D.** Genome browser sashimi plots from IGV viewer showing extended 3'ends as labeled above from two representative histone genes *Hist1h1b* and *His1h2bp* in each of *Zcchc8*$^{+/+}$, *Zcchc8*$^{+/-}$, *Zcchc8*$^{-/-}$ transcriptomes, respectively. **E** and **F.** Sashimi plots for two coiled-coil domain containing cilia genes *Ccdc89* and *Ccdc182,* respectively by genotype. For *Ccdc89*, in addition to accumulation of reads beyond the mature 3'-end in the *Zcchc8*$^{-/-}$ transcriptome, there is also an increase in discontinuous upstream of gene 5'end reads that resemble so-called PROMoter uPstream Transcripts (PROMPTs). 72

Chapter One

Introduction

Telomerase is limiting

Telomeres are the repetitive DNA sequences capping the ends of chromosomes ensuring end-replication does not interfere with genome stability. Telomerase functions to synthesize new telomere repeats. It has two essential components: the telomerase reverse transcriptase (TERT) catalytic component and a telomerase RNA (TR) template (*1, 2*). Telomere length is exquisitely sensitive to perturbations in these two essential components, and disruption in the availability of either component leads to telomere shortening (*3, 4*). Loss of TERT and TR lead to haploinsufficiency for telomere shortening in yeast, mice, and humans (*5-10*). In humans 6 of 13 disease genes including either TR itself or TR-related proteins have be found to be mutated in short telomere syndromes (STS) (*11*). However, despite advances in genome-wide sequencing and targeted next-gen sequencing of these known genes, a significant fraction of sequenced families in the Johns Hopkins short telomere registry have low TR levels without identifiable mutations in known genes. Causative mutations in these families could be evasive, deep intronic and other non-coding changes affecting known genes that are yet to be uncovered. Alternatively, they could be mutations in novel genes with roles in TR RNA biogenesis. Screens for genes critical for telomere length regulation in yeast suggest a significant fraction of telomere length regulating genes (one-third) are involved in RNA metabolism, leading us to hypothesize that unsolved human families with STS especially those with aberrant TR levels have mutations in genes involved in RNA biogenesis critical for stability, processing, or maturation (Figure 1) (*12, 13*).

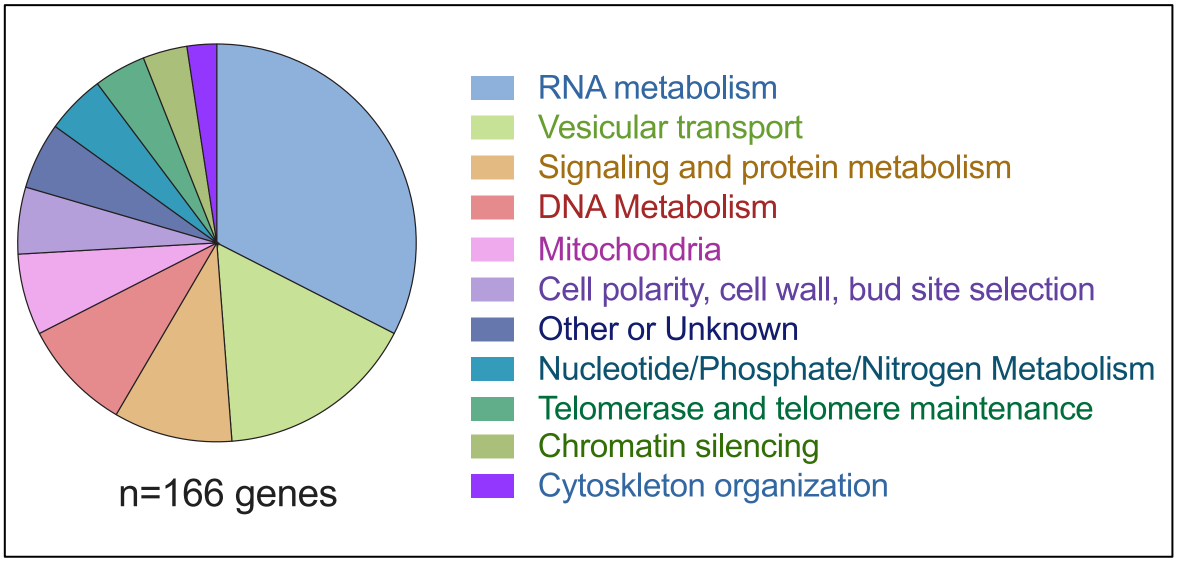


Figure 1. Genes identified as regulators of telomere length in yeast. Unique genes were identified as having roles in telomere length regulation after review among three independent gene deletion, haploid, and essential gene screens for telomere length maintenance in *Saccharomyces cerevisiae (12-14)*. A total of 166 genes were grouped into 10 categories and the frequency of genes in each category is represented here in this pie chart.

Mutations in TR-related proteins cause dose-dependent clinical manifestations

Mutations in TR-related proteins cause clinical manifestations of short telomere syndromes specifically due to perturbations in mature TR levels that lead to telomere shortening. It is the telomere length and not telomerase RNA levels that influence disease as has been shown in mice. Loss of TR in both $TR^{-/-}$ and $TR^{+/-}$ mice causes telomere length shortening across generations leading to decreased survival in late-generation mice and tissue renewal defects as shown in intestinal epithelium, testes seminiferous tubules, bone marrow, and peripheral blood. These mice show anticipation of short telomere disease as seen in humans, with more accelerated shortening in $TR^{-/-}$ compared to $TR^{+/-}$

mice that reflects a dose-dependent relationship between TR levels and disease severity (*6, 15*). This dose-dependent relationship with disease severity also exists for TR-related proteins mutated in humans. Individuals with heterozygous loss-of-function mutations in nuclear assembly factor 1, *NAF1*, and poly(A)-specific ribonuclease, *PARN*, are haploinsufficient for TR and develop short telomere syndromes (*16-18*). Biallelic hypomorphic *PARN* mutations also cause a more severe disease, Hoyeraal-Hreidarsson syndrome (HHS), with clinical manifestations including severe bone marrow failure, immunodeficiency, developmental defects and cerebellar hypoplasia (Table 1) (*17, 19*). No biallelic *NAF1* mutations have been identified yet to date, but in this work we engineered mice with a null *NAF1* allele to examine the molecular consequences driving disease upon *NAF1* loss. We found TR levels were disrupted in $Naf1^{+/-}$ mice but had intact pseudouridylation whereas $Naf1^{-/-}$ mice were embryonic lethal, underscoring the exquisite sensitivity of TR levels to this dose-dependent mechanism for disease severity. Further, we found that heterozygous loss of function mutation of a novel gene, *ZCCHC8*, in familial short telomere syndrome disrupts TR 3’ end processing and maturation. We go on to show that complete loss of ZCCHC8 in knockout mice causes a more severe neurodegenerative phenotype than ZCCHC8 heterozygous loss in which mice develop only a telomere phenotype. The fact that heterozygous loss of TR-related proteins cause only short telomere disease support the idea that non-telomeric functions of those proteins are only perturbed upon biallelic loss causing more severe phenotypes such as intellectual disability or lethality.

Table 1. Dose-dependent manifestations of mutations in TR and TR proteins.

Gene	Heterozygous mutations	Biallelic mutations
TR	Short telomere syndrome	Dyskeratosis congenita
NAF1	Short telomere syndrome	None reported
PARN	Short telomere syndrome	Dyskeratosis congenital and HHS syndrome
ZCCHC8	Short telomere syndrome	? Neurodevelopmental defect

TR sequence, structure, and ribonucleoproteins are highly divergent across species

Although the requirement for a reverse transcriptase and RNA template essential for telomere elongation in humans, yeast, and many other lower organism was discovered and first characterized in cilia (*2, 20*), TR is highly divergent in sequence and structure across phylogenetic kingdoms which complicates the utility of identifying homologous TR ribonucleoproteins bound to these RNAs in lower organisms specifically as candidate genes to inform vertebrate TR biogenesis. The distinct secondary structures in each species require a unique collection of ribonucleoproteins for maturation and stability which may contribute distinct 3'-end processing mechanisms. Cilia have the shortest telomerase RNA. tTER in *T. thermophilia* is only 159 bp and bound by p65 which interacts with p75, p50, p45, and p19 proteins for stability and Teb1 recruitment for telomere lengthening. TER1 and TLC1 in *S. pombe* and *S. cerevisiae* respectively are much longer (>1,200 nucleotides) requiring "long range" base-pairing in 3 stems and additional ribonucleoproteins for stability. Both TLC1 and TER1 are bound by Sm proteins and Est1; however, TLC1 is additionally bound by Pop1, Pop6, and Pop7, the Ku 70/80 complex, and Est2 indirectly bound with Est3. Human TR ribonucleoproteins are altogether different than cilia and yeast as it has evolved from yeast to contain a H/ACA domain which is bound by Dyskerin complex at the 3'end (*21, 22*). Interestingly although lower organism TRs do not have H/ACA motifs, there are small nucleolar (sno) and small Cajal body associated (sca) RNAs in yeast that do have H/ACA motifs which have informed vertebrate TR biogenesis. For example, the only human proteins with yeast homologs that have been identified in short telomere syndrome are proteins essential to H/ACA biogenesis including *DKC1, NAF1,* and *PARN*. Studies of these proteins in yeast sno- and sca- RNA maturation has shed light on mechanism for vertebrate telomerase H/ACA stability and processing at the 3'end (*21*). However despite these differences and insights from yeast, many questions remain regarding the mechanisms for vertebrate TR processing at the 3'end and its interaction, if any, with the 5' end. Do the distinct ribonucleoproteins in each species dictate a distinct 3'end processing mechanism and are there any similarities in 3'end processing of the lower organism TR with the highly divergent vertebrate TR?

Table 2. TR processing characteristics in cilia, yeast, and humans.

Organism	**RNA, Length (nt)**	**Transcription**	**Ribonucleoproteins**	**5' end**	**3' end**
T. thermophilia	*tTER*, 159	RNA Pol III	p65, p75, p19, p45, p50, Teb1	No capping	Polyuridine tail, no processing
S. pombe	*TER1*, 1231	RNA Pol II	Lsm proteins, Est1	trimethylguanosine capping by Tgs1	Spliceosome cleavage
S. cerevisiae	*TLC1*, 1301	RNA Pol II	Sm proteins, Est1, Est2, Est3, Ku70, Ku80	trimethylguanosine cap by Tgs1	NRD1 complex and RNA exosome
H. sapiens	*hTR*, 451	RNA Pol II	Dyskerin, NAF1, NOP10, NHP2, GAR1, TCAB1	trimethylguanosine cap by TGS1	PARN, ZCCHC8 and the RNA exosome?

End processing of 3’ extended TR has evolved as a step towards maturation

In addition to these differences in structure and binding proteins, the 3’end processing mechanisms and involved proteins among cilia, yeast, and humans also vary widely. Cilia require no 3’end processing as tTER ends in a polyuridine tail that requires no further maturation. In yeast, *S. pombe* utilizes the spliceosome for endonucleolytic cleavage of nascent TER1 to its mature form whereas *S. cerevisiae* coopts the RNA exosome for TLC1 exonuclease trimming mediated by Nrd1-Nab3-Sen1 complex that is limited by the RNA bound Sm proteins (Table 2) (*23-25*). A poyladenylated 3’extended TLC1 exists and comprises a small percentage (<10%) of the total RNA; however polyadenylation is not required for maintaining mature levels which suggests the polyadenylated form is not a mature TR precursor and is dispensable for telomere length maintenance in *S. cerevisae* (*26*). Similarly, a 3’ extended polyadenylated form of human telomerase RNA (hTR) also exists, but in contrast to *S. cerevisae*, disruption of polyadenylation due to mutations in the polyadenylase *PARN* decreases the total fraction of mature 451 nucleotide hTR leading to telomere length shortening and subsequently disease in humans (Table 2) (*16, 17, 27*). To date, the mechanism of hTR processing to the mature form with a shorter 3’extended polyadenylated intermediate processed by PARN remains poorly understood. The mechanism of PARN dependent maturation of the 3’ end of RNAs, in particular snoRNA H/ACAs, was first described in U2OS cancer cells (*28*) and 3’ RACE deep sequencing first used to identify a population of oligoadenylated human TR in HEK293T cells (*29*) before the these findings and methods were adapted to describe PARN dependent TR maturation in induced pluripotent stem cells from patients with biallelic *PARN* mutations and STS (*17*). Nascent TR forms greater than 1000 nucleotides exist

(*30*) but how and whether these are intermediates are processed to the mature forms has remained elusive. In this work, we propose ZCCHC8 targets these long 3’ extended nascent TR forms to the RNA exosome for 3’ exonucleolytic trimming to mature or near-mature TR forms that can become polyadenylated and matured by PARN. We use genetic studies in familial short telomere syndromes and functional studies utilizing 3’end sequencing technologies in cell lines to provide insight into the mechanisms of 3’end TR processing by establishing a role for the RNA exosome in maturation as in *S.* cerevisiae via a nuclear exosome targeting complex.

Vertebrate TR transcription could be coupled to the RNA exosome and dependent on localization

Unlike sno- and sca- H/ACAs which are excised from introns of mRNA genes, TR is transcribed from its own promoter and locus and because it is matured starting from its own transcriptional unit, it is plausible that the 5’ and 3’ end processing have distinct features from other H/ACAs and are possibly even coupled by yet to be determined mechanisms (*31*). Human and yeast TR biogenesis starts with transcription by RNA pol II. Both mouse and human TR contain promoter features typical of other pol II mRNAs including a TATA box and consensus CCAAT sequence (*32*). Transcription includes addition of a 5’ monomethylated cap of TR that is subsequently trimethylated by TGS1 which has been shown to be localized at the Cajal body (*33-35*). The methylated cap serves a number of roles for RNAs during transcription including interaction with the Cap Binding Complex (CBC) which consist of the cap proteins CBP20 and CBP80. The CBC is recruited by RNA pol II to the 5’ cap of nascent RNA to support protection, transcription, termination, and 3'end formation of nascent RNAs. It carries out these functions, including 3'end formation, by recruiting other RNA processing factors including the snRNA trafficking protein PHAX and the Nuclear RNA EXosome Targeting (NEXT) complex via the zinc finger RNA exosome mediating protein, ZC3H18 (*36-38*). The PHAX snRNA export protein has also been shown to bind TR within the nucleus, but how it contributes to maturation has not been explored (*39*). The CBC complex has recently been implicated in TR processing as extended nascent TR precursors accumulated upon transient knockdown of CBP80 in HeLa cell lines (*40*). However, little evidence exists to date supporting a mechanism by which the CBC contributes to TR maturation specifically through coupling to the NEXT complex and RNA exosome at the 3'end. The RNA exosome has also been linked to TR degradation via the TRF4/AIR2/MTR4 polyadenylation (TRAMP) complex (*30, 40*). Studies suggest the TRAMP complex acts to degrade RNA via the RNA exosome in the nucleoli. The NEXT complex and DIS3 RNA exosome exonuclease are localized in the nucleoplasm and its role in TR maturation has not been explored (*37*). Overall, subcellular localization of RNA exosome components could be a determinant of maturation via the NEXT complex and DIS3 verses degradation in the nucleoli via the TRAMP complex. In this thesis we show that the NEXT complex component ZCCHC8 is required for TR 3'end formation, and it is plausible that this 3' end maturation via the NEXT complex is dependent on coupling to CBC at the 5'-end at the trimethyl cap. Testing of this hypothesis and mechanism will be the subject of future investigation.

The data in this thesis support this model for vertebrate TR 3'end processing

In summary, we utilized using genome-wide linkage in a family with short telomere syndrome to identify a new gene, the NEXT complex component ZCCHC8, which targets human TR to the RNA exosome for maturation. We propose a model in which nascent TR is transcribed by RNA polymerase II to an extended precursor as long as or greater than 1000 nucleotides which requires exonucleolytic trimming to the most common mature 451 nucleotide form. The extended nascent TR is recruited to the RNA exosome for processing to the shorter mature form for by the NEXT complex in the nucleoplasm via ZCCHC8 binding, potentially by coupling to the CBC at the 5' trimethylguanosine cap. Near-mature polyadenylated nascent forms (less than 10-15 nucleotides extended) are subsequently processed by PARN to the mature form. The Dyskerin complex binds TR at the H/ACA domain to provide stabilization as NAF1 traffics TR to the Cajal body where it is exchanged with GAR1 for subsequent recruitment of TERT and trafficking to the telomere for telomere elongation (Figure 2).

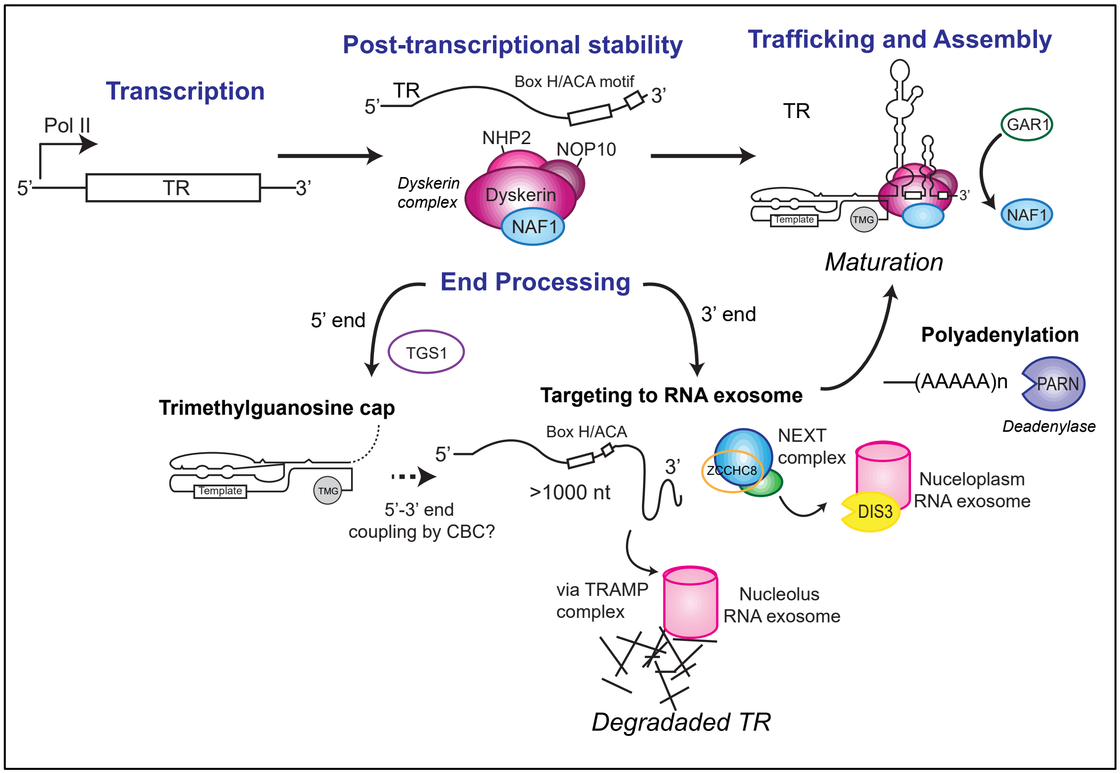


Figure 2. Model for hTR biogenesis and processing. hTR is transcribed from its own promoter (right angle arrow) by RNA polymerase II (pol II). Post-transcriptional stability is provided by binding of the Dyskerin complex to the TR box H/ACA motif. The Dyskerin complex also includes NAF1, NHP2, and NOP10. End processing at the 5' and 3' end is shown (large curved arrows). TGS1 trimethylates the 5' trimehthylguanosine cap (TMG) which may be coupled to the cap binding complex (CBC) to assist in 3' end processing (dashed arrow). Extended nascent TR transcripts as long as or greater than 1000 nucleotides (nt) are either targeted for maturation via binding by ZCCHC8 and the NEXT complex to the RNA exosome bound by the exonuclease DIS3 in the nucleoplasm or targeted for degradation via the TRAMP complex to the RNA exosome in the nucleolus. After trimming by the RNA exosome, shorter adenylated nascent TR forms (abbreviated with line and short polA tail) are processed further by the deadenylase PARN to the most common mature TR form that is 451 nt and bound by the Dyskerin complex. NAF1 traffics TR to the Cajal body where it is exchanged with GAR1.

Chapter Two

Haploinsufficiency of the H/ACA biogenesis factor *NAF1* selectively disrupts telomere length maintenance while sparing ribosomal RNA functions

Dustin L. Gable[1,2,3], Susan E. Stanley[1,2,3], Thomas M. Carlile[4], Vidya Sagar Hanumanthu[2], Jonathan K. Alder[1], Wendy V. Gilbert,[4] Mary Armanios[2,3,5,6]

Medical Scientist Training Program[1]
Department of Oncology[2]
Telomere Center[3]
McKusick-Nathans Institute of Genetic Medicine[5]
Sidney Kimmel Comprehensive Cancer Center[6]
Johns Hopkins University School of Medicine
Baltimore, MD 21287

Department of Biology, Massachusetts Institute of Technology[4], Cambridge, MA

Current affiliations: SES is currently a resident at Massachusetts General Hospital.

INTRODUCTION

Telomerase is a ribonucleoprotein consisting of both an essential reverse transcriptase (TERT) and integral RNA component (TR) that are necessary for telomere elongation during replication (*2, 20*). Additional proteins are important for telomere maintenance, recruitment of telomerase for elongation, and telomerase RNA biogenesis and trafficking, and 13 of these genes have previously been linked to short telomere syndromes including ribonucleoproteins of the dyskerin complex dyskerin itself, NOP10, and NHP2 (*41*). Almost a decade ago, the discovery was made that inherited short telomeres can manifest as pulmonary fibrosis with the discovery of germline mutations in *TERT* and *TR* in familial disease(*9*). More recently short telomeres and mutations in *TERT* and have also been found in severe cases of COPD (*42*).

Recently, using a candidate gene approach as previously described, we have uncovered heterozygous, frameshift mutations in nuclear assembly factor one, *NAF1,* in both a familial and sporadic case of pulmonary fibrosis and short telomeres (*18*). NAF1 is a ribonucleoprotein that associates with the dyskerin complex and has been shown to be important for box H/ACA RNA stability (*43, 44*). Small nucleolar (sno-) and small Cajal body (sca-) RNAs containing H/ACA box domains have multiple functions including one of the most frequntly described roles in guiding site-specific pseudouridylation of ribosomal, spliceosomal, and messenger RNAs, an enzymatic process shown in some cases to be important for ribosomal integrity and translation at the ribosome. (*45, 46*). Similar to these sno- and scaRNAs, TR contains an H/ACA box domain previously shown to be essential for its stability and localization in the Cajal bodies, and mutations in this domain are known causes of disease (*8, 47-53*).

We studied null *NAF1* mutations in families with low TR, short telomeres, and pulmonary fibrosis. Lymphoblasts from cases having the familial *NAF1* mutation had reduced NAF1 levels, were haploinsufficent for TR, and had telomere lengths measuring below the 1st percentile. Additionally, TR levels were reduced upon NAF1 knockdown in HeLa cells and a homozygous knock-in model of the familial *NAF1* mutation in HCT116 cells dramatically decreases TR levels (*18*). Here, we generated mice heterozygous for a null mutation in *Naf1* and show that haploinsufficiency of Naf1 in these mice leads to a marked reduction in TR levels, but no reduction in pseudouridylation at sites (Ψ) in ribosomal RNA (rRNA) despite a trending reduction in a select group of guide H/ACA RNAs targeting these sites. Further, we show despite the decreases in TR, snoRNAs, and scaRNA levels, these mice have no clinical pathologies and NAF1 null mice were embryonic lethal likely due to disruption pseudouridylation and rRNA stability. These findings suggests that the telomerase is sensitive to NAF1 dosage and the phenotype associated with loss of NAF1 is a telomere syndrome and does not cause pathologies due to disruptions in rRNA biogenesis.

RESULTS

Decreases in mammalian NAF1 result in decreased TR levels

Naf1p is essential in yeast (Fatica, Capozzo, Dez); however, its essential role in mammalian systems has yet to be described. What is known has been limited to its function in maintaining small H/ACA RNA levels in yeast and HeLa cells. Given our interest in the regulation of telomerase RNA which also contains a box H/ACA, and to ascertain its essential role in mammals, we used CRISPR/Cas9 technology to create null alleles in the mouse. We first tested this model in mouse embryonic fibroblasts (MEFs). We designed an sgRNA and DNA oligo for homology directed recombination (HDR) to target and edit a locus in exon 1 of *Naf1,* that would simultaneously introduce a stop codon and alter an endogenous SacII restriction cut site to NaeI. A restriction digest and gel electrophoresis were used to screen for clones with editing, followed by Sanger sequencing. Digest of the 838 base pair (bp) into a 599 bp and 239 bp band indicated cutting. Cutting with SacII restriction digest to two bands indicated the presence of a WT allele while cutting with NaeI indicated the presence of a correctly edited, null allele (Figure 3a). A CRISPR/Cas9 vector containing the sgRNA was transfected with the HDR oligo into a MEF cell line. 14 clones grew after a serial dilution of transfected cells. Because there were so few, we screened for CRISPR cutting directly by Sanger sequencing. 2 of 14 clones (14%) were edited. Both clones contained 3 mutant alleles, one in frame and two frameshift mutations. Aneuploidy is a common alteration in MEFs, which was confirmed in our cell line by karyotype analysis (data not shown, reference). Notably, there were no complete *Naf1* knockouts. Western blotting using an antibody generated that is specific to mouse NAF1 (mNAF1) showed over an 80% decrease in NAF1 levels in both the mutant cell lines (Figure 3b, c). A similar percentage decrease was also seen in TR levels as measured in each by qRT-PCR. (Figure 3d). Overall, this evidence suggests that *Naf1* could be an essential gene and that decreases in the levels of NAF1 levels result in proportional decreases in TR levels.

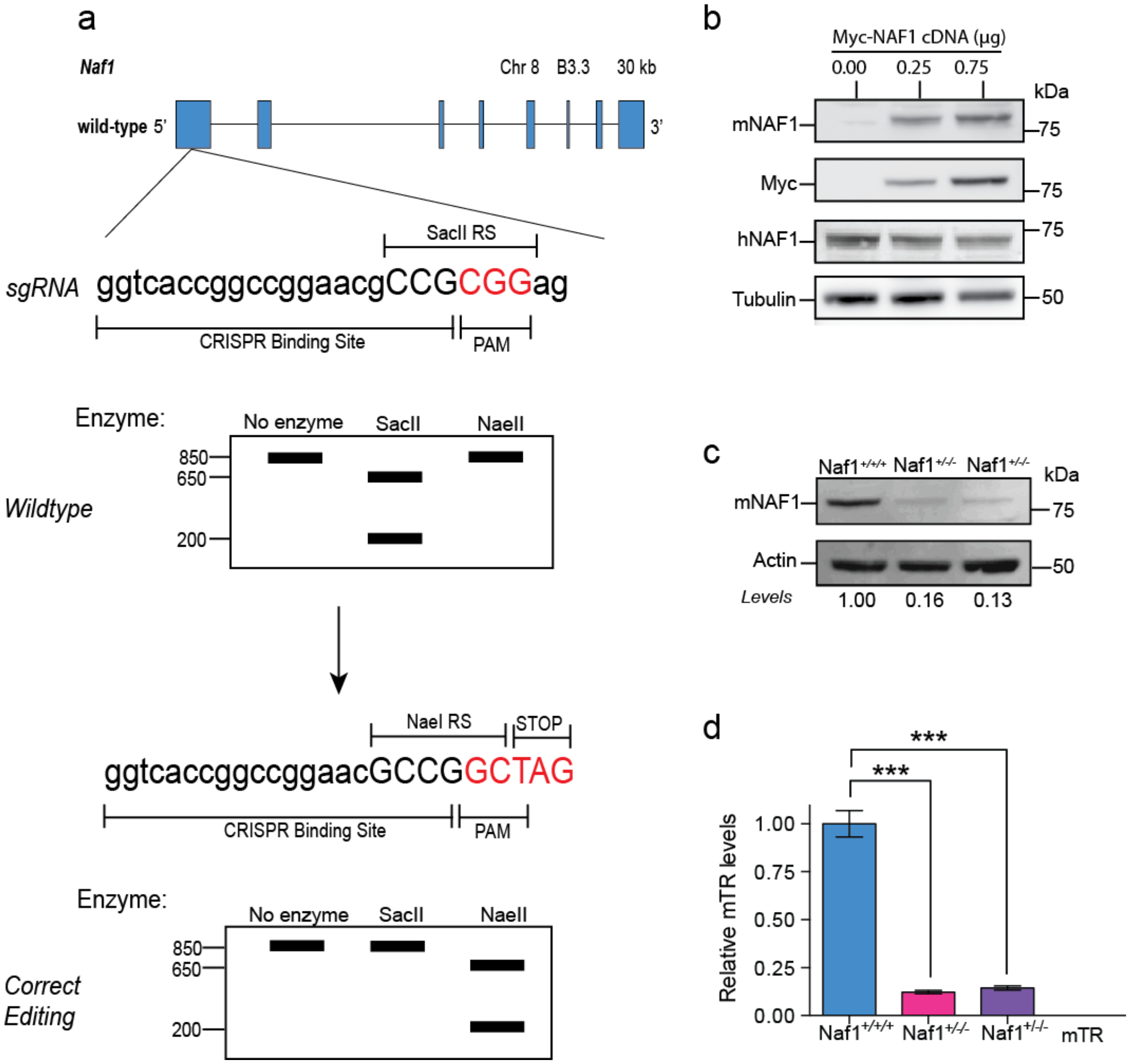


Figure 3. TR levels are decreased in mouse embryonic fibroblasts with decreased NAF1 levels. a. Schema of mouse wild-type *Naf1* locus and CRISPR/Cas9 editing strategy to knockout *Naf1*. A sgRNA was designed to target exon 1 and introduce a nonsense mutation via homology directed recombination that would also alter the native *SacII* restriction site to a correctly edited *NaeII* restriction site that. Restriction digest of the target region PCR product with either enzyme followed by gel electrophoresis could be used to detect the presence of a wildtype or correctly edited allele. **b.** Immunoblot of lysate from human embryonic kidney (HEK) 293FT cells transfected with Myc-tagged Naf1 cDNA-containing plasmid at shown concentrations. Immunoblot for human NAF1 (hNAF1) and mouse NAF1 and Myc shows specificity of the murine NAF1 antibody. **c.** Immunoblot for NAF1 in CRISPR edited mouse embryonic fibroblasts. **d.** TR measured by qRT-PCR in *Naf1*$^{+/+/+}$ (blue) MEFs compared to two different *Naf1*$^{+/-/-}$ mutant MEF cell lines (pink and purple) each containing one allele with an in frame mutation and two with

framseshift mutations. Levels were relative to hypoxanthine-guanine phosphoribosyltransferase (Hprt). For both *Naf1*$^{+/+/+}$ versus *Naf1*$^{+/-/-}$ comparisons, P <0.001, n=3 technical replicates.

NAF1 is an essential gene in mammals

Studies in the MEF cell line point towards *Naf1* being an essential gene in mammalian systems. To definitively address this observation, we decided to knockout *Naf1* in a mouse model using the same CRISPR/Cas9 strategy used in the MEFs (Figure 3b). The sgRNA, Cas9 protein, and HDR oligo were injected into single cell zygotes of C57BL/6J mice, which were then implanted into surrogate mothers. 6 mice were screened for editing using restriction digest with SacII and NaeI enzymes. Cutting only occurred with the SacII enzyme suggesting there was no correct editing with the HDR oligo. 2 mice (2 and 4) contained only wildtype alleles similar to the control wildtype MEFs, 1 mouse was heterozygous for a wildtype and mutant allele, and 3 mice had no cutting with either enzyme indicating two mutant alleles (Figure 4a). Sanger sequencing via TOPO cloning was carried out on pups containing two mutant alleles. There were 6 mutant alleles among the 3 mice, and mouse 1 contained a predicted frameshift null mutation (c.320_641del322bp, p.G108Cfs*77) which was chosen to isolate and breed *Naf1* null mice (Figure 4 a, b). This mouse was backcrossed to a wildtype mouse to generate a *Naf1*$^{+/-}$ heterozygous mouse which had half the levels of NAF1 compared to *Naf1*$^{+/+}$ as measured by western blotting (Figure 4c). *Naf1*$^{+/-}$ heterozygous pups were then interbred across multiple generations to look for *Naf1*$^{-/-}$ mice, none of which were observed (>200 screened, $P < 0.001$, Fisher's exact test). Embryos were isolated as early as day E8.5-E12.5 and genotyped, and again no *Naf1*$^{-/-}$ embryos were observed (n=32, P<0.01). These findings allowed us to conclude that *Naf1* is an essential gene in mice. Additionally, *TR*$^{-/-}$

mice are viable suggesting that the essential role in NAF1 lies beyond telomere homeostasis, and could involve RNA biogenesis.

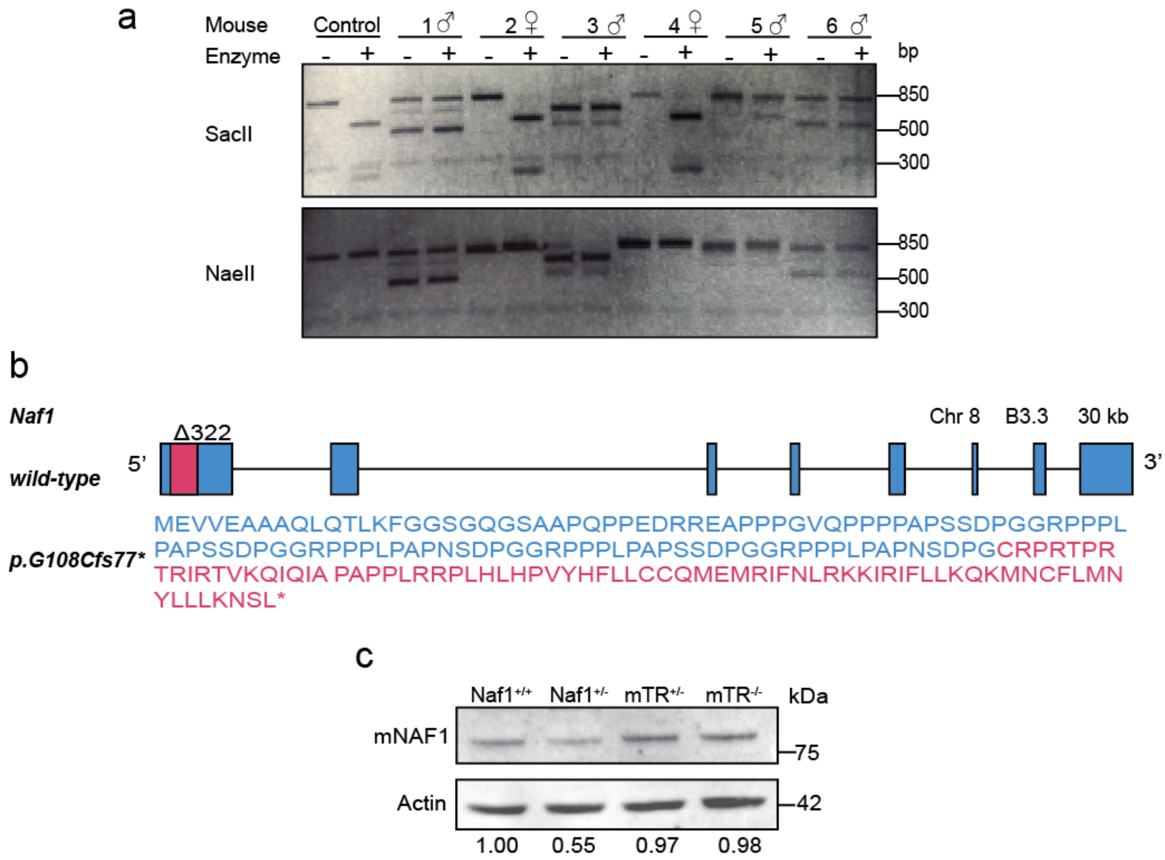


Figure 4. ***Naf1*$^{+/-}$ mice were created using CRISPR/Cas9 editing.** a. Genotyping using the *SacII* and *NaeI* restriction digest strategy (Figure 3a) of C57BL/6J mice created from injection of Cas9 protein, sgRNA targeting mouse *Naf1* exon 1, and an oligo for homology directed recombination into single cell zygotes compared to control wildtype MEFs (n=6 mice). b. Schema of wild-type *Naf1* locus and CRISPR/Cas9-induced 322-bp deletion (Δ322) in exon 1 and the amino acid sequence created by the frameshift mutation created (blue=wildtype, pink=mutant). c. Immunoblot for NAF1 in adult skin fibroblasts relative to Actin.

NAF1 is necessary for telomerase RNA and noncoding RNA stability

Given the proportional decrease in TR levels observed upon CRISPR knockdown of NAF1 levels in MEFs, we first measured TR levels in *Naf1*$^{+/-}$ mice to investigate whether heterozygous loss of NAF1 results in haploinsufficency for TR. *Naf1*$^{+/-}$ mice had half the levels of TR compared to *Naf1*$^{+/+}$ mice (61% by qRT-PCR and 51% by northern compared to controls, n=3, P 0.05). Notably, TR levels in *Naf1*$^{+/-}$ mice were also significantly increased compared to *TR*$^{+/-}$ mice whose levels were 44% by qRT-PCR and 41% by northern compared to *Naf1*$^{+/+}$ mice (Figure 5a,b,c). This difference between *Naf1*$^{+/-}$ and *TR*$^{+/-}$ mice is likely a reflection of heterogeneous TR levels among mice, among cell types, and within cells as has been shown in previous studies measuring TR levels in mouse testes and telomerase activity in various mouse tissues (*54*). However, this difference could also implicate NAF1 in TR 3' end processing. A polyadenylated TR precursor was recently described that can accumulate when there is a processing defect involving proteins involved maturation or degradation including *PARN* (*17*); however, we measured polyadenylated levels of TR in *Naf1*$^{+/-}$ mice by priming RNA with oligo(dT) and normalizing TR levels measured by qRT-PCR to that of RNA primed by random hexamer which would measure total TR (mature+precursor) and found no increase in polyadenylated precursor TR (data not shown). This evidence points towards the biological variability as an explanation for the difference in TR levels between *Naf1*$^{+/-}$ mice and *TR*$^{+/-}$ mice. Given the significant decrease in TR levels and known role in H/ACA snoRNP complex, we were interested in the effect of decreased NAF1 levels on other RNA types. We measured 12 different H/ACA box, C/D box, long noncoding RNA (lncRNA), small nuclear RNA (snRNA), and mRNA levels by qRT-PCR in *Naf1*$^{+/-}$ mice and compared that to *Naf1*$^{+/+}$ mice. Among 8 H/ACA RNAs, levels were on average 63%

(range 46 to 74%) of controls in *Naf1*$^{+/-}$ mice (n=3 to 8 mice), 3 were significantly decreased which

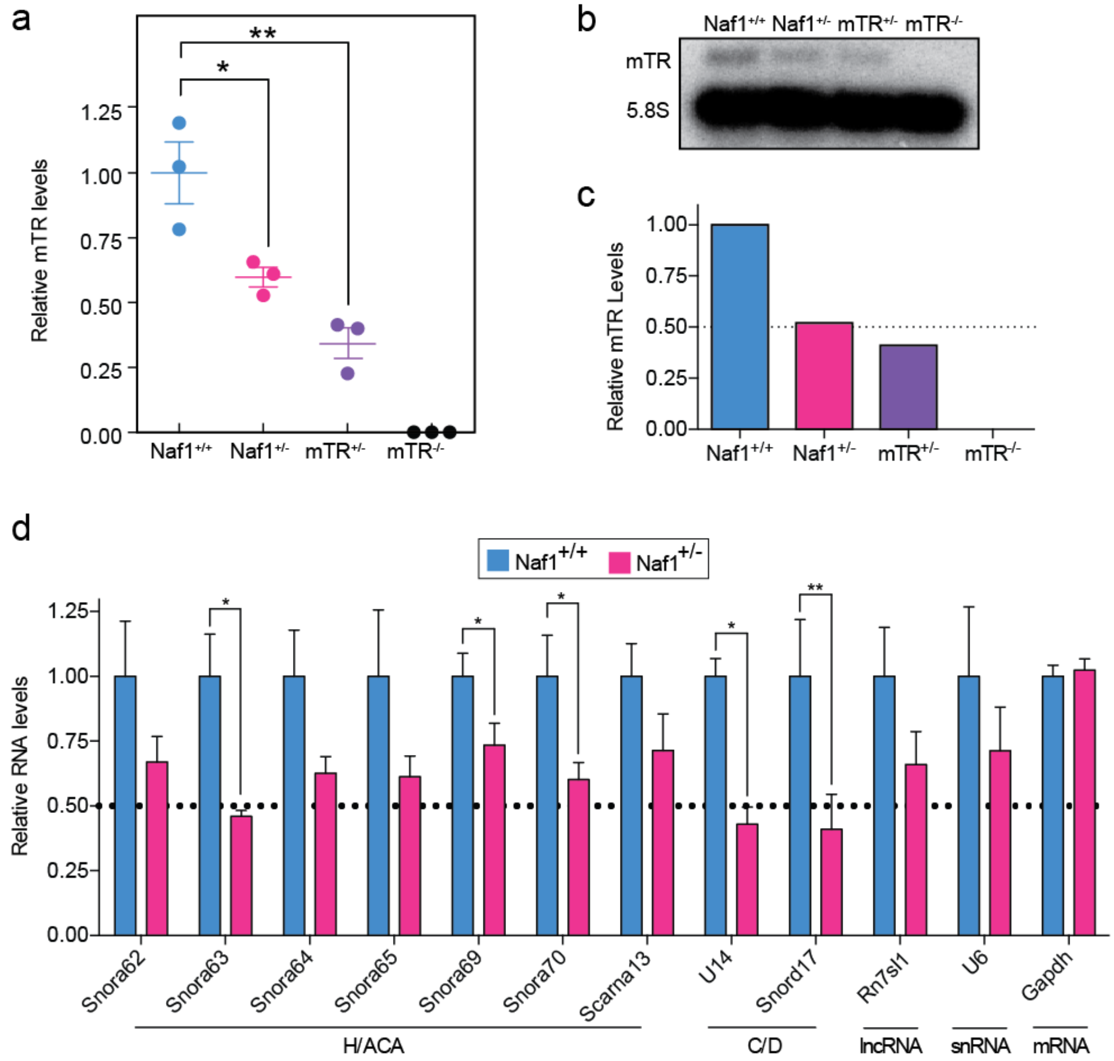


Figure 5. RNA Levels in Naf1$^{+/-}$ mice. a. TR measured by qRT-PCR in Naf1$^{+/+}$, Naf1$^{+/-}$, mTR$^{+/-}$, and mTR$^{-/-}$ mice (total spleen RNA, n = 3 mice per group) and normalized to hypoxanthine-guanine phosphoribosyltransferase (Hprt). b-c. Northern blot for mTR (mouse TR) normalized to 5.8S (b) with quantifications of Naf1$^{+/+}$(blue), Naf1$^{+/-}$ (pink), mTR$^{+/-}$ (purple), and mTR$^{-/-}$ below (c). d. H/ACA, C/D, splicing (snRNA), long noncoding (lnRNA) and messenger (mRNA) RNA levels measured by qRT-PCR (n = 3 to 8 mice per group, total spleen RNA) and normalized to Hprt in Naf1$^{+/+}$(blue), Naf1$^{+/-}$ (pink). Error bars represent SEM. **P < 0.01 and ***P < 0.001 (Student's t test).

included included Snora64, Snora68, and Snora70 (n=8, $P < 0.05$). U14 and Snord17 C/D RNAs were significantly decreased at 43 and 41% of controls (n=3, $P < 0.05$ and 0.01 respectively). Both a lncRNA and snRNA had trending, but non-significant decreases in TR levels (66 and 71% decreases respectively) and GAPDH mRNA was not decreased (n=3) (Figure 5d). Overall, it appears that *all* non-coding RNAs had significant or trending decreases in RNA levels, including the lncRNA Rn7sl1 which is 299 base pairs and functions in the cytoplasm in association with the signal recognition particle ribonucleoprotein. This is surprising given that previous reports in yeast and HeLa cells suggest NAF1 is required specifically for H/ACA RNA stability (yeast and HeLa reports).

Naf1$^{+/-}$ **mice do not have detectable generational telomere shortening**

Haploinsufficiency of telomerase RNA in *Naf1*$^{+/-}$ mice raises the questions as to whether telomere shortening is detectable. To measure telomere shortening we interbred *Naf1*$^{+/-}$ mice across 4 generations and measured telomere length by Flow-FISH (Figure 6a). No significant generational shortening was detected across all four generations of *Naf1*$^{+/-}$ mice. Specifically the difference in the geometric mean ratio between *Naf1*$^{+/-}$G4 and *Naf1*$^{+/+}$G1 mice was only 0.03 ± 0.02 (n=4 mice per generation, P=0.2647, Student's t test, Figure 6b). The inability to detect shortening is not surprising given the long and heterogenous telomere length distribution in C57BL/6J mice. In contrast, CAST/EiJ mice have shorter and homogenous telomere length and significant shortening in *TERT*$^{+/-}$ mice can be detected even in 6 generations accompanied with decreasing testes and spleen weights (*55*). To corroborate the conclusions that the telomere length in the *Naf1*$^{+/-}$

C57BL/6J mice was too long and heterogenous to have a clinical consequence and detect shortening across generations, the testes and spleens of each mouse were weighed prior to telomere length measurement, and neither testes nor spleen weights decreased across each generation of *Naf1*$^{+/-}$ mice (Figure 6c)

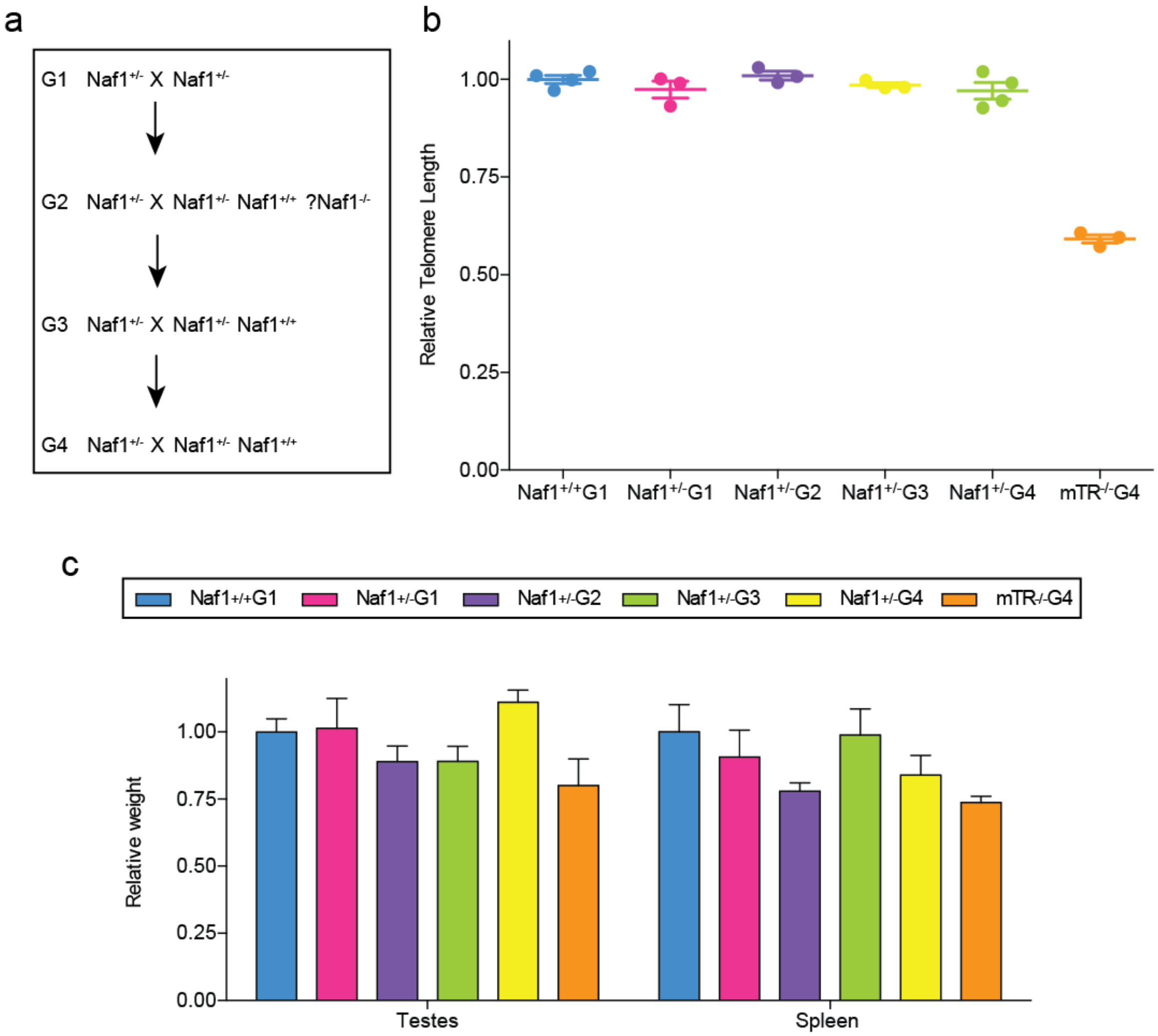


Figure 6. Telomere length in Naf1$^{+/-}$ mice. a. Generational breeding strategy for generating *Naf1*$^{+/-}$G1, *Naf1*$^{+/-}$G2, *Naf1*$^{+/-}$G3, and *Naf1*$^{+/-}$G4 mice for telomere length measurement. **b.** Relative telomere length in lymphocytes isolated from spleens of *Naf1*$^{+/+}$ G1 (blue), *Naf1*$^{+/-}$ G1 (pink), *Naf1*$^{+/-}$ G2 (purple), *Naf1*$^{+/-}$ G3 (yellow), *Naf1*$^{+/-}$ G3 (green), and *mTR*$^{-/-}$G4 (orange) mice measured by measured by flow cytometry and

fluorescence in situ hybridization (flow-FISH) (n=3 mice per group). **c.** Relative weight of testes and spleens isolated from mice whose telomere lengths were measured in *b* (n=3 mice/ group).

rRNA pseudouridylation in *Naf1*$^{+/-}$ mice is intact.

H/ACAs function to guide targeted pseudouridylation at specific RNA uridine residues which is understood to be important for RNA integrity and function (*21, 46*). Over 133 rRNA sites are predicted to be pseudouridylated, and given the decreases in H/ACA RNA levels in *Naf1*$^{+/-}$ mice we wanted to query modifications at rRNA sites using a recently developed high throughput, quantitative assay known as Pseudo-seq (*46*).We identified 107 pseudouridines by homology to human rRNA sites, and of those, 82 sites met the quality cut off for analysis. The aggregate pseudouridine (Ψ) signal across all the Ψs detected in *Naf1*$^{+/-}$ mice was not significantly different from wild-type (96.7%, P=0.2, unpaired *t*-test, Figure 7a). Examining each individual rRNA site revealed a similar Pseudo-seq signal and near equal correlation of Pseudo-seq signal at each of 82 sites in *Naf1*$^{+/-}$ mice compared to wildtype (R^2=0.954, P<0.001) (Figure 7b). Additionally, because the rRNA sites analyzed by Pseudo-seq were identified based off conservation of mouse to human, we show targeted complementary binding of 6 mouse H/ACA RNAs at 9 rRNA positions to verify that our rRNA sites analyzed were true H/ACA target sites (Figure 8a-i). Importantly, the levels of these H/ACA RNAs were measured by qRT-PCR, and despite downward and significant decreases in RNA levels, pseudouridylation at these sites was not compromised suggesting that pseudouridylation is vital for rRNA function.

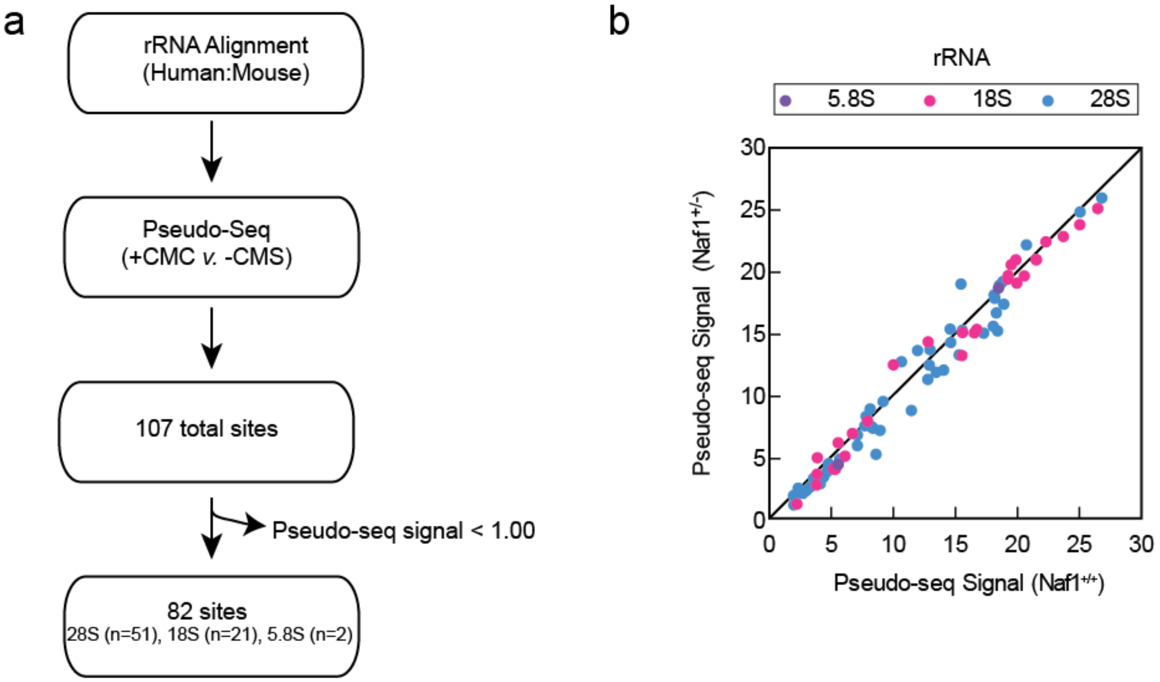


Figure 7. Pseudouridylation is intact in *Naf1*$^{+/-}$ mice. a. Alignment and Pseudo-seq strategy for identification of 82 pseudouridylation sites in mouse rRNA. b. Scatter plot of average Pseudo-seq signal ratio in *Naf1*$^{+/+}$ (n=4) versus *Naf1*$^{+/-}$ (n=3) mice graphed for 82 sites in three rRNA subunits.

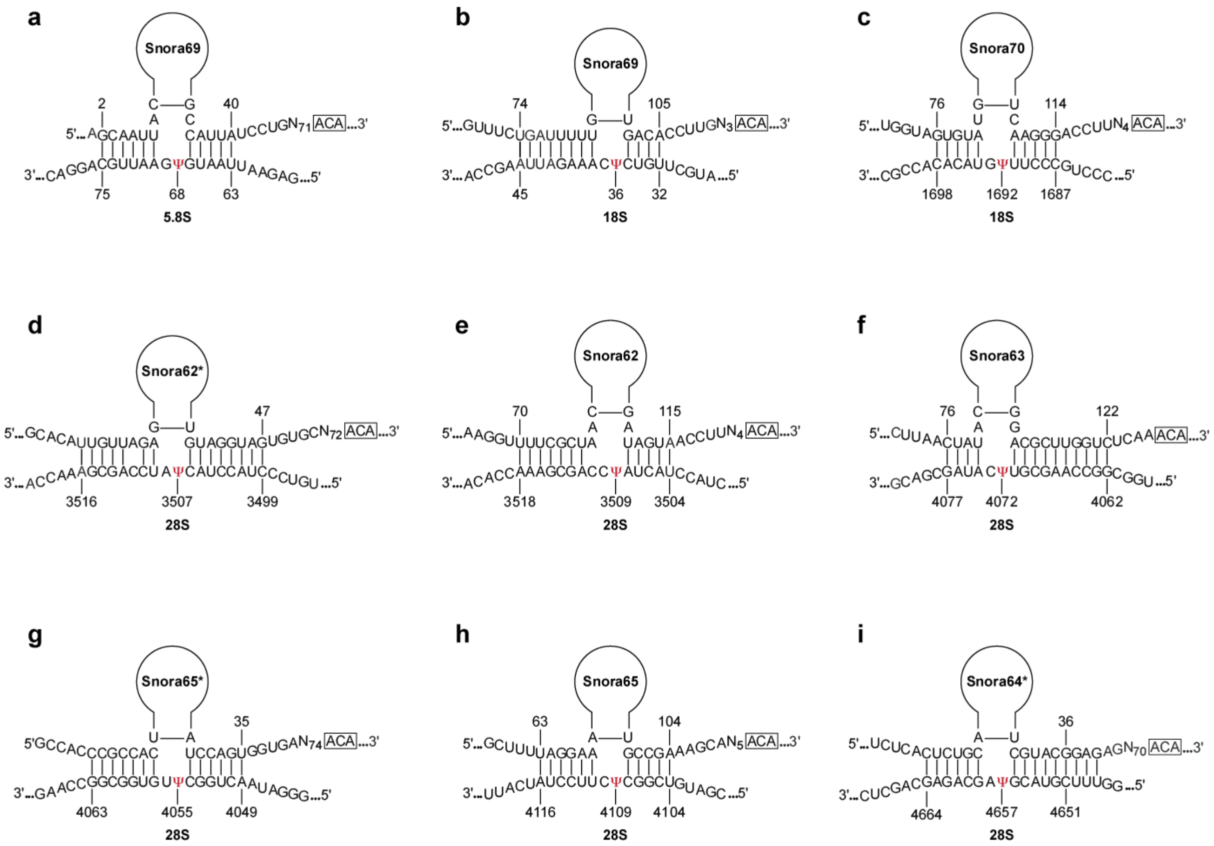


Figure 8. Schema of mouse box H/ACA snoRNAs quantified showing the complementary guide sequence and the rRNA pseudouridine (Ψ) target site (red). a-i. The six snoRNA quantified in Fig 2.3d are shown here with their complementary sequence targeting 9 rRNA sites (shown in panels a-i). Each rRNA respective subunit is shown below. Eight of these rRNA sites were detected by Pseudo-seq as annotated in Table 1. In panels *d, g,* and *i* (designated by *), the murine RNA reference sequence was incomplete at the 5' end, and sequence complementarity was derived from the genomic sequence based on build GRCm38/mm1.

Naf1$^{+/-}$ mice do not have features of ribosomopathies.

To further explore potential consequences of decreased H/ACA RNA levels, we wanted to clinically assess *Naf1*$^{+/-}$ mice. We hypothesized that if there were a clinical defect associated with the decreased levels, it could resemble ribosomapthies given the importance of H/ACA RNA levels in modifying and maintaining ribosomal integrity. In humans, ribosomopathies are a hetergenous group of rare diseases with autosomal dominant or recessive inheritance caused by mutations in over 20 different ribonucleoproteins. Syndromes caused by these disorders commonly include craniofacial abnormalities and hematologic defects such as bone marrow failure, anemias, and cancer (Armistead, Triggs-Raine, *FEBS Letters*, 2014). Over 11 different types of mouse models currently exist for ribosomopathies which were created by engineering missense, null, or splicing mutations in a number of ribonucleoproteins. Clinical phenotypes of these mice are also heterogeneous and include but not limited to hematopoietic, skeletal, and craniofacial defects, and growth retardation (Terzian and Box, *PLOS,* 2015). We phenotyped first-generation *Naf1*$^{+/-}$ mice by performing necropsies including gross and histological tissue analysis of the head, abdomen (stomach, small intestine, colon, liver, spleen) and male reproductive organs (testes, epididymis, vas deferens, seminal vesicles, prostate, and penis) along with complete blood counts (white blood cells, neutrophils, lymphocytes, monocytes, red blood cells, hemoglobin, hematocrit, and platelets). No significant defects were observed (n=6 *Naf1*$^{+/+}$*,* n=5 *Naf1*$^{+/-}$). Overall, the intact pseudouridylation and clinically healthy phenotype of *Naf1*$^{+/-}$ mice suggests that the molecular defects associated with decreased H/ACA RNA levels are sufficient to compensate maintain cellular homeostasis within the mouse. Further, given that *TR*$^{-/-}$ mice are viable, the essential nature of *Naf1* may likely be explained by its role in RNA biogenesis.

DISCUSSION

Here we report findings that *NAF1* is an essential gene in mice with only a telomere phenotype without defects in pseudouridylation in heterozygous state. Prior to these studies, pseudouridylation in mice secondary to defects in the dyskerin complex had only been studied in transgenic point-mutant catalase-inactive dyskerin embryonic stem cells, fibroblasts, and lymphocytes They report decreases in pseudouridylation in cell lines as measured by thin layer chromatography but with no analysis of systemic defects. (*56, 57*). Further, dyskerin is X-linked and male knockout mice are embryonic lethal and clinical assessment of mouse models was limited to a dyskerin hypomorph and low resolution pseudouridylation analysis of tissues and cells(*58*) (*59*). *Naf1* mice here provide dan autosomal model of disease to cleanly study the consequences of 100% v 50% v 0% of dyskerin complex member NAF1. Dyskerin levels were intact in humans with heterozygous frameshift mutations in NAF1 and previous western blotting not shown here for dyskerin in *Naf1* heterozygous null mice shows intact levels pointing to a NAF1 specific defect in these mice. Pseudouridylation of rRNA as measured by high throughput sequencing is intact in *Naf1*$^{+/-}$ mice despite decreases in H/ACA RNAs as shown here and there is no clinical evidence of a ribosompathy as described in a multi-system phenotypic analysis. Whether or not pseudouridylation is defective as a consequence of dyskerin loss was not examined here, but our studies highlight H/ACA RNA levels cannot be a surrogate for function of pseudouridylation in other models. Futher, high-throughput pseudouridylation analysis through Pseudo-seq will be a standard for pseudouridylation compared to TLC in assessing these questions in the future. We would hypothesize that some baseline unction of dyskerin will be sufficient to maintain adequate pseudouridylation in humans with dyskerin mutations as complete loss of pseudouridylation is lethal. Overall, *Naf1*$^{+/-}$ mice haploinsufficient for TR model patients with heterozygous loss of TR and point toward a purely telomeric effect in those individuals.

Decreases in C/D along with H/ACA RNAs in *Naf1*$^{+/-}$ mice was surprising. These data could support a role for *NAF1* in stabilization of non H/ACA small RNAs in mice through mechanisms not yet understood. One possibility to help explain this observation is that nascent pre-processed C/D RNAs could have H/ACA domains in the 3'ends that are removed in processing upon maturation. Alternatively mouse *NAF1* in mice could identify C/D domains through a different mechanism. Mouse and human NAF1 proteins are only 82% identical with poor conservation at the N terminus. Functional analysis at this protein region could shed light onto whether or not additional regulatory functions of NAF1 beyond H/ACA RNA stability and targeting exist. C/D RNAs are involved in methylation of rRNA and methyl-seq of NAF1 mutant mice and MEF cell lines with 30% of NAF1 could shed light as to wether C/D RNA function is compromised (*21*). However given our findings of intact pseudouridylation in heterozygous NAF1 mice, we would hypothesize methylation to also be intact. A side by side pseudo-seq and methyl-seq experiement of MEF cell lines could offer important clarity. Alternatively, our observations may reflect important differences in *in vitro* versus *in vivo* systems.

Finally, these data highlight the sensitivity of telomerase RNA dosage in cell lines in that just 50% loss of NAF1 in mice is sufficient to cause disrupt telomere homeostasis whereas 50% loss of additional H/ACA is tolerated. This difference between TR and other H/ACAs is likely due to the low abundance of TR available to telomere elongation in the cell. Point mutations in human short telomere syndromes disrupting TR secondary structure are sufficient for destabilization of TR but even point mutations not disrupting base-pairing cause disease highlight the exquisite sensitivity of TR to perturbations normal homeostasis (*60*). Other sno and sca-H/ACAs are highly abundant on the order of magnitude of 1000 fold high experession than TR such that loss of 50% may be tolerated. Pseudouridylation of rRNA is essential for life in one generation as we have shown here and cells may have adapted a large preserve to compensate when H/ACA levels are perturbed or compromised. Additionally, H/ACA RNAs have different regulatory mechanisms than TR being most commonly excised from introns of messenger RNAs whereas TR is a single exon gene with its own promoter (*21*). It is possible that loss of NAF1 has less effect on H/ACA function beyond stability given differences in regulatory function but we know loss of TR stability is sufficient for disease. Overall, telomerase RNA is tightly regulated and disruption of NAF1 is sufficient to cause telomere mediated disease in the context of normal pseudouridylation and rRNA integrity.

METHODS

Mice

Mice were housed in the Johns Hopkins University School of Medicine East Baltimore campus, and all procedures were approved by its Institutional Animal Care and Use Committee. All strains were maintained on the C57BL/6J background. $TR^{+/-}$ and $TR^{-/-}$ mice were generated and maintained as described (Blasco reference).

Generation of *Naf1* null mouse embryonic fibroblasts

Targeting the Naf1 locus was performed using CRISPR/Cas9 (reference). A synthetic single-guide RNA (sgRNA) target sequence (5′-GGTCACCGGCCGGAACGCCG-3′) was designed to target exon 1 of Naf1(NC_000074.5) using the Optimized CRISPR Design tool (http://crispr.mit.edu/). This oligonucleotide was cloned into a pX459 plasmid (gift from F. Zhang; Addgene, no. 48139) (reference). Prior to editing, karyotyping in $Trf2^{Fl/Fl}$ MEFs was performed showing aneuploidy (Celli reference). 150,000 cells were plated in a 6-well dish one day prior to transfection and then transfected with 2 μg of pX459 plasmid using Lipofectamine 3000 (Invitrogen). After 24 hours, cells were selected under 2 mg/mL of puromycin for 72 hours. Serial dilutions of selected cells were plated into 96-well plates and single colonies were grown out and screened for locus editing by Sanger sequencing, n=14 colonies (F- 5’ gctgcagacgctcaagttc 3’, R- 5’ aaaggaagcgggaactccta 3’).

Generation of *Naf1* null mice

Targeting the Naf1 locus was performed using the same CRISPR design described for mouse embryonic fibroblasts. However the pX459 was appended with a T7 promoter, and the sgRNA was in vitro transcribed and purified. The sgRNA and Cas9 (TriLink BioTechnologies) were co-injected into C57BL/6J zygotes (Johns Hopkins Transgenic Core). Pups (n = 6) were screened for locus editing by Sanger sequencing: 5′-GCTGCAGACGCTCAAGTTC-3′ (forward) and 5′-AAAGGAAGCGGAACTCCTA-3′ (reverse) (838-bp wild-type allele). Mice were subsequently genotyped using this primer pair.

Mouse skin fibroblast derivation and mouse phenotyping. Adult mouse skin fibroblasts were derived from wild-type, Naf1$^{+/-}$, TR$^{+/-}$, and TR$^{-/-}$ mice (7-18 weeks). Briefly, ear punches were washed in PBS containing penicillin streptomycin fungizone, minced in Hank's Balanced Salt Solution (Gibco), and then digested with Collagenase B (Roche) and 0.25% Trypsin-EDTA (Gibco) for 20 minutes each at 37°C. The tissue was then washed and plated in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplied with fetal bovine serum, glutamine, penicillin and streptomycin at standard concentrations. Timed matings were performed as previously described with vaginal plugs indicating E0.5 (9). E8.5-E12.5 embryos were dissected in PBS and DNA was extracted using the Gentra Purgene Mouse Tail kit (Qiagen). Phenotyping was performed on age- and sex matched (11-15 weeks, n=5/group) first generation Naf1+/- mice. Complete blood counts were measured using the ProCyte Hematology Analyzer (IDEXX Laboratories). Necropsy and tissue histology analyses were performed in a blinded study by a veterinary pathologist and included brain, liver, gastrointestinal tract, spleen, and testes gross and microscopic analyses.

Naf1 antibody generation. A polyclonal rabbit antibody against mouse Naf1 (Naf1 394) was generated using a synthetic peptide containing amino acids 394-408 (NP_001157036.1) with the addition of an N-terminal cysteine (C-QDRGSDASWKNDQE) (ProSci) as described (70). Antibody specificity was verified by plasmid transfection of Myc-tagged Naf1 (NM_001163564.1) in HEK-293FT (followed by immunoblotting using Myc and Naf1 394 antibodies.

Immunoblotting studies

Immunoblotting was performed as described (60). Briefly, cultured cells were lysed using radioimmunoprecipitation assay buffer supplemented with a protease inhibitor cocktail (Roche). Proteins were resolved with SDS–polyacrylamide gel electrophoresis, and after transfer to a nitrocellulose membrane were blocked in Odyssey blocking buffer (LI-COR). The following primary antibodies were used: mouse Naf1 (rabbit, Naf1 394, 1:250; Prosci) and actin loading control (mouse, ab8226, 1:2000; Abcam). Blots were visualized visualized by horseradish peroxidase– linked antibody (rabbit, 7074 and mouse, 7076 1:10,000; Cell Signaling Technology) and enhanced chemiluminescent substrate (Thermo Scientific).

TR qRT-PCR

Total RNA was extracted with the RNeasy Mini Kit (Qiagen) as recommended except that deoxyribonuclease digestion was performed using twice the volume and for an extended period of 1 hour. Random hexamer-primed cDNA was synthesized using the SuperScript III Reverse Transcriptase kit (Invitrogen). Quantitative PCR of mTR was carried out as described (reference), respectively. Minor modifications included using freeze-thaw cycles to lyse cell pellets in lysis buffer and homogenization of total mouse spleens using the Bullet Blender (Next Advance). For each experiment, the side by-side comparisons were made on simultaneously prepared cDNA.

TR northern blot. Northern blot was performed as previously outlined (62) with minor modifications. Briefly, 25 µg of total RNA was resolved on a 1.5% agarose-formaldehyde gel. RNA was vacuum transferred to HybondTM-N+ (GE Healthcare, Amersham) nylon membrane in 10x SSC for 1 hour followed by UV crosslinking. Pre-hybridization was performed in Church buffer [500 mM sodium phosphate (pH 7.2), 1 mM EDTA, 1% BSA, 7% SDS] for 1 h at 65ºC. DNA probes were labeled with [α-32P] dCTP by random priming. Membranes were hybridized with mTR and 5.8S probes overnight in Church buffer at 65ºC. The membrane was washed with two quick rinses at 65ºC in 2x SSC solution, followed by five rinses in 2x SSC and 0.1% SDS, at room temperature (5 minutes each), and then for 30 minutes in 0.1x SSC and 0.1% SDS. Blots were exposed to a phosphor screen and visualized on Storm 825 system (GE Healthcare).

The mTR probe, which included the 397 bp mTR sequence plus 142 bp of flanking sequences from the genomic locus, was excised from plasmid DNA (Bluescript KS+ JHU1145, gift from Carol Greider, Johns Hopkins University) by Not1/Sal1 restriction digest (New England BioLabs) and gel-purified. The human and mouse 5.8S sequences are 100% identical and we used the same probe amplified from mouse gDNA to detect

5.8S as a loading control for mTR. The probe was generated using this primer pair: 5'-GACTCTTAGCGGTGGATCAC-3' (forward), 5' GACGCTCAGACAGGCGTAG-3' (reverse) (152 bp).

H/ACA RNA identification and quantification by qRT-PCR. We quantified seven mouse H/ACA RNAs and identified them based on ≥ 80% sequence identity with their human homologs. Sequence identity between mouse and human RNAs were identified using BLASTn using these mouse sequences: Snora62 (NR_002902.2), Snora63 (ENSMUST00000082448), Snora64 (NR_002897.1), Snora65 (NR_002898.2), Snora69 (NR_002900.1), Snora70 (NR_002899.1), and Scarna13 (NR_028576.1). Target sites in mouse rRNAs [(5.8S (NR_003280.2), 18S (NR_003278.3), and 28S (NR_003279.1)] were extrapolated based on the known target sites in human rRNA using the snoRNABase Version 3 Database (71). They were then verified manually by the presence of a complementary and ACA sequence. For qRT-PCR analysis, standards were generated using plasmid DNA containing each RNA sequence. These were cloned from mouse total spleen cDNA using random hexamer priming. The details of the qRT-PCR procedure were as described for mTR. Primer sequences and PCR conditions are available upon request.

Telomere length measurement. Telomere length was measured by flow-FISH (57). The flow-FISH method includes extensive controls, including MESF (molecules of equivalent soluble fluorochrome) beads and cow thymocyte internal control (57). Telomere length data were plotted as the relative geometric mean intensity of *Naf1*$^{+/+}$G1 mice.

Pseudo-Seq library preparation. Pseudoridine site quantification was performed on total RNA extracted from mouse spleens by Pseudo-Seq as previously described with minor modifications (39). The total and site-specific pseudouridine quantification was performed on the same samples. Briefly, libraries were prepared by fragmenting 5 μg total RNA in 10 mM ZnOAc at 60ºC for 20 minutes, followed by CMC modification and reversal. RNA fragments (60-80 nt) were gel purified, and prepared for reverse transcription (RT) by ligation of a pre-adenylated adapter to the 3′ ends. Following RT, truncated cDNAs were gel purified, circularized with circLigase (Epicentre) at 60ºC for 6h and PCR amplified. Libraries were sequenced on an Illumina HiSeq2000.

Identification of mouse rRNA pseudouridines and pseudouridine quantification. Mouse rRNA pseudouridines were identified in two steps. First, conservation of modification between human and mouse was assumed. The positions in mouse rRNAs corresponding to known human rRNA pseudouridines (39, 71) were determined by aligning human (18S: X03205, 28S: U13369 nts 7935-12969, 5.8S: U13369 nts 6623-6779) and mouse rRNAs (18S: NR_003278.3, 28S: NR_003279.1, 5.8S: NR_003280.2) with Clustal Omega (72). Second, additional pseudouridylation events in mouse rRNA were identified by analysis of Pseudo-seq libraries with the following peak calling parameters: a peak cutoff of 1.0, and reproducibility in >80% of libraries. Reads were processed and mapped to the mm10 transcriptome (UCSC, downloaded May 13, 2015). Peak values were calculated as follows: For a given Ψ, the reads whose 5’ ends map to each position in a 51 nt window centered at the Ψ were normalized to the total number of reads in the window. This normalization was performed for both the +CMC and –CMC libraries. The normalized reads at each position in the window in the –CMC library were subtracted from the normalized reads at that position in the +CMC library. Finally, each position in the +CMC and –CMC calculation was scaled by the window size. Only those sites with a peak value of greater than 1.0 in 6 of 7 +/- CMC library pairs were considered for further analysis. A total of 107 Ψ sites were identified in this analysis; these are listed in Table 1 with the detected Ψ sites (n=82) annotated. The variance in measurement of site-specific pseurouridine ratios was compared within a given genotype, Naf1+/+ (n=4 assayed) and Naf1+/- (n=3 mice assayed), and showed an intra-group ratio range of 2% to 10%. The raw Pseudo-seq data were deposited in NCBI's GEO database (submission ID 17772667, Accession ID GSE78063).

Statistical analyses

Statistical comparisons were made with GraphPad Prism software, and all P values are two-sided. Student's t test was used for comparison of means, and P values less than 0.05 were considered significant.

Chapter Three

ZCCHC8 is mutated in pulmonary fibrosis

Dustin L. Gable[1,2,3], Valeriya Gaysinskaya[2,3], Susan E. Stanley[1,2,3], Elizabeth W. Pugh[4],
Kara M. Schenk[5], Murat O. Arcasoy[6], Liliana Florea[4], Mary Armanios[2,3,4,7]

Medical Scientist Training Program[1]
Department of Oncology[2]
Telomere Center[3]
McKusick-Nathans Institute of Genetic Medicine[4]
Osler Medical Housestaff Training Program[5]
Sidney Kimmel Comprehensive Cancer Center[7]
Johns Hopkins University School of Medicine
Baltimore, MD 21287

Department of Medicine[6]
Duke University School of Medicine
Durham, NC 27708

Current affiliations: SES is currently a resident at Massachusetts General Hospital.

INTRODUCTION

Telomerase is a specialized ribonucleoprotein that synthesizes telomere repeats (*1, 20*). It has two core components: TERT, the telomerase reverse transcriptase, and TR (also known as TERC), the telomerase RNA component, which provides the template for telomere addition (*2, 31, 61*). In humans, mutant telomerase genes have their most common manifestation in a degenerative age-related disease known as idiopathic pulmonary fibrosis (IPF) (*9, 62*). Mutant telomerase genes are the most common cause of autosomal dominant IPF, and a subset of IPF patients show systemic features of a short telomere syndrome including bone marrow failure (*63*). Among the IPF families with identifiable mutations, half carry mutations in TERT or TR (*9, 64*). The other half carry mutations in one of five other telomere maintenance genes, RTEL1, PARN, NAF1, TINF2, and DKC1 (*16, 18, 65-67*). Four of the seven heretofore identified telomere maintenance genes linked to familial IPF affect TR itself, or its de-adenylation, trafficking or biogenesis as is the case for PARN, NAF1 and DKC1, respectively (*17, 18, 68*).Fifty to 70% of familial IPF patients have no identifiable cause and, within this subset, there are individuals with obvious features of a short telomere syndrome and unexplained low TR levels (*18*). One hurdle to gene discovery in these families is that candidate genes have been exhausted. Moreover, combining gene discovery efforts across families is problematic because there is extreme locus heterogeneity with each of the recently discovered genes accounting for only 1-3% of uncharacterized cases (*16, 18*). There is also the additional challenge of assigning affected status clinically since IPF is generally a disease of older adults and, once diagnosed, it usually has a rapidly progressive course.

Here, we used molecular and clinical phenotyping with genome wide linkage analysis to identify a novel regulator of TR maturation that is mutated in familial pulmonary fibrosis. Human TR has its own promoter and is transcribed by RNA polymerase II. Vertebrate TR shares a 3'end box H/ACA motif with a subset of other non-coding RNAs, but relative to these and other mRNAs, it has very low abundance (*18, 47, 69*). This scarcity makes telomere length maintenance highly vulnerable to even minor reductions in TR dosage and function (*3, 4*). In human families, even hypomorphic *TR* mutations cause autosomal dominant IPF with genetic anticipation (*9, 70*). The regulation of vertebrate TR transcription, processing and assembly with TERT remain incompletely understood in part because TR sequence is divergent and organisms have adapted distinct mechanisms for post-transcriptional processing and 3'-end maturation (*69*) (*71*). For example, T. thermophila TR is transcribed by RNA polymerase III that relies on runs of Ts for transcriptional termination (*2*). S. cerevisiae TR (known as TLC1) relies on a 3'end Sm-binding complex in addition to the Nrd1-Nab3-Sen1 complex for proper 3' end processing by the RNA exosome (*24, 25*). S. pombe on the other hand uses a modified spliceosome-dependent cleavage mechanism as the initial 3'end processing step (*25*). Vertebrate TR is transcribed as a longer transcript and while its processing has been suggested to be linked to the RNA exosome (*30, 40*), the determinant components for its maturation are not known. Here, we report mutant ZCCHC8, the zinc finger CCHC-type domain containing 8 protein, as a cause of TR insufficiency and familial IPF. ZCCHC8 was identified as a component of the Nuclear Exosome Targeting Complex (NEXT) (*37*), which has been implicated in targeting a subset of RNAs to the nuclear RNA exosome, but has not been linked to TR maturation. We show that ZCCHC8 is required for nascent TR processing.

RESULTS

Genome wide linkage analysis identifies mutant ZCCHC8 in a family with autosomal dominant pulmonary fibrosis

We identified an adult with IPF who showed classic short telomere syndrome features including bone marrow failure. His family history was consistent with autosomal dominant pulmonary fibrosis (Figure 9A). He had low TR levels in lymphoblastoid cell lines by qRT-PCR, similar to a patient with mutant DKC1 (Figure 9B). Whole genome sequencing identified ~3600 rare variants that deviated from the reference genome, but none of them fell in known TR-related genes. To identify the disease-causing variant, we recruited 13 of his relatives, but found no clinically affected individuals (Figure 9A). This led us to assess their preclinical status by measuring TR levels and we found the proband's two sons similarly had 50% lower TR levels compared to unrelated healthy controls as well as marry-ins from the same family (Figure 9B). The other remaining relatives had intact TR levels (Figure 9B). Telomere length by flow cytometry and FISH (flowFISH) confirmed the two children were pre-clinically affected since they both had abnormally short telomere length, well below the first age-adjusted percentile, and had shorter telomere length than the proband (Figure 9C). This finding is similar to the successive telomere shortening seen in families with TERT and TR mutations which underlies telomere-mediated genetic anticipation (*63, 70, 72*). The molecular studies supported assigning pre-clinical affected status to the two pre-symptomatic individuals, in addition to

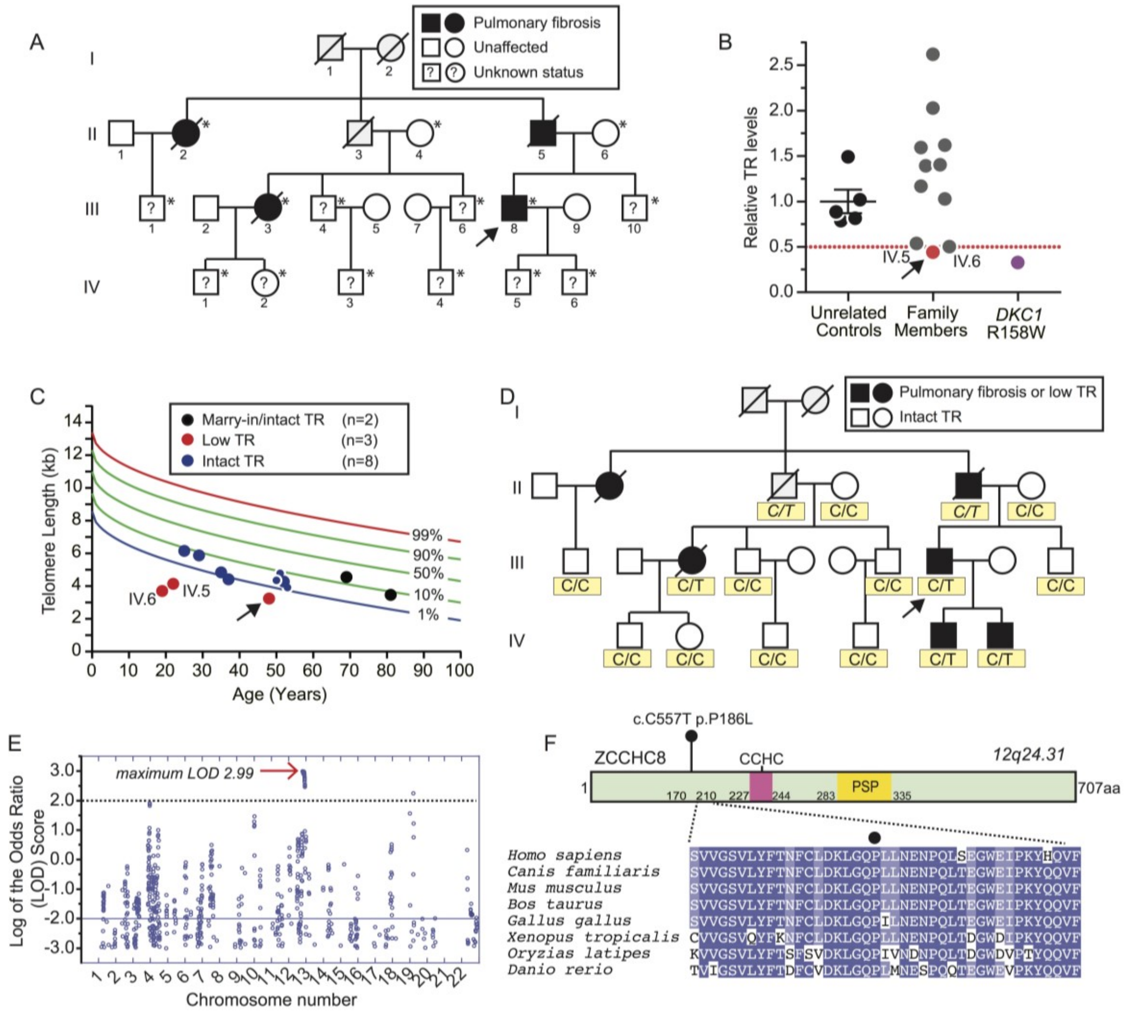


Figure 9. Linkage analysis identifies the *ZCCHC8* locus containing a novel mutation that segregates with familial pulmonary fibrosis and telomerase RNA (*TR*) insufficiency. A. Pedigree with proband (designated by arrow) and relatives who had pulmonary fibrosis are indicated by the shaded symbols (Key). The '?' refers to individuals whose clinical status was uncertain at the time of clinical assessment and the individuals shaded in gray had an unknown cause of death. *Refers to individuals for whom DNA was available. B. *TR* levels measured by quantitative real time PCR (qRT-PCR) in lymphoblastoid cell lines (LCLs). Arrow refers to proband (red dot) and pedigree identifiers refer to A (proband's sons). *TR* levels

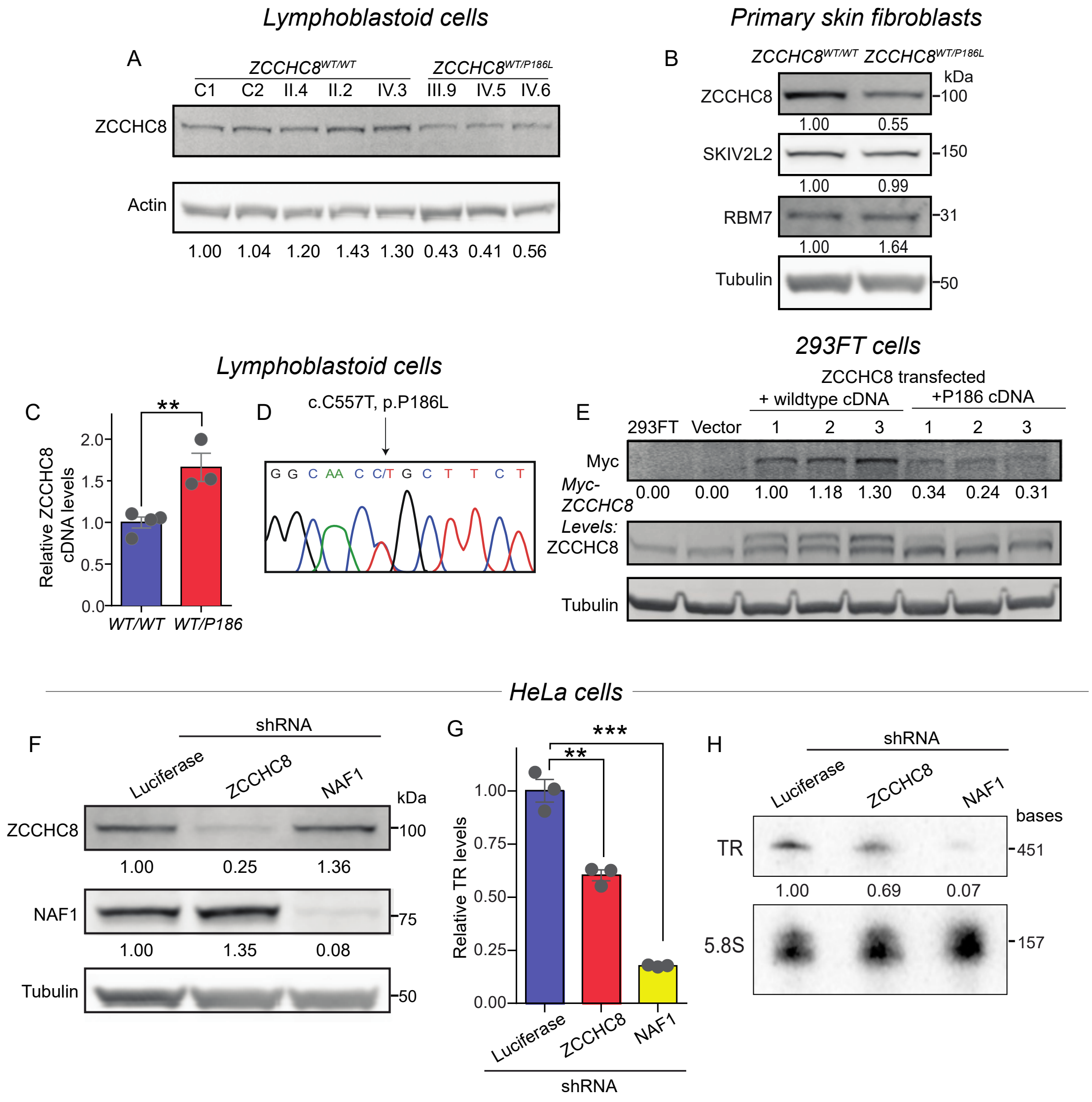
from a *DKC1* mutation carrier is a positive control. The data represent a mean of three experiments, each from independent RNA isolations. C. Telogram shows age-adjusted lymphocyte telomere length by flow cytometry and fluorescence in situ hybridization (flowFISH) in the proband (arrow) and his family. Pedigree designations are as in A. The telogram is based on clinically validated dataset from 192 controls. D. Phenotype assignments as shown in key that were used in genome wide linkage analysis based on the presence of pulmonary fibrosis as well as *TR* levels (B) and telomere length measurement (C). The *ZCCHC8* single nucleotide variant for the P186 change is also shown below each pedigree symbol and italicized genotypes refer to obligate carriers. E. Log of the odds ratio score across the autosomal chromosomes calculated from 15 individuals identified by * in (A). The linkage peak on chromosome 12 is denoted by the arrow. F. Novel protein coding mutation identified p.P186L falls in an unstructured domain of *ZCCHC8* with alignment across eight vertebrate species shown. The darker shading denotes more conserved identity conserved residue identity. CCHC refers to Zinc-knuckle domain; PSP refers to proline-rich domain.

the proband, and the remaining family members, who had intact TR levels, were designated as unaffected (Figure 9D). Using these assignments, we performed genome wide linkage analysis and identified a single 17.3 Mb linkage peak on chromosome 12 (1073260-1246120) which had a maximum log of the odds ratio (LOD) score of 2.99 (Figure 9E). Within this interval, there were three rare coding and no rare splice-altering variants. Only one of these variants was novel; it fell in an RNA metabolism gene, ZCCHC8. The variant predicted a missense mutation c.C557T, p.P186L and was absent in more than 140,000 individuals (genome Aggregation Database, Exome Variant Server, 1000 Genomes, dbSNP) (Figure 9F). Within the family, the *ZCCHC8*^{P186L} variant segregated with the pulmonary fibrosis and low TR-short telomere phenotypes, including in obligate carriers (Figure 9D). P186 falls in a highly conserved but unstructured domain of ZCCHC8 (Figure 9F). We screened 42 genetically uncharacterized families with pulmonary fibrosis and other short telomere syndrome phenotypes by targeted sequencing of *ZCCHC8*, but we did not identify additional variants, suggesting these mutations are rare in uncharacterized cases (1 of 43, 2%). This finding is consistent with the locus heterogeneity of short telomere syndromes.

***ZCCHC8*^{P186L} is a loss-of-function mutation**

To characterize the significance of *ZCCHC8*^{P186L}, we tested whether it affected protein integrity and found mutation carriers had nearly 50% lower ZCCHC8 protein levels in lymphoblastoid cells compared to controls and non-carriers in the same family (P=0.038, Mann-Whitney U-test, Figure 10A and antibody validation not shown). The lower ZCCHC8 protein levels were also seen in proband's primary skin fibroblasts (Figure 10B). Notably, the decrease was specific to ZCCHC8 as we saw no significant change in the levels of the other two NEXT complex proteins, RBM7 and SKIV2L2 (Figure 10B and antibody validation not shown). We tested whether the low protein expression was due to decreased mRNA or protein stability and found that ZCCHC8 mRNA levels were intact in mutation carriers (mean 1.6-fold ± 0.2 s.e.m relative to controls, Figure 10C). Moreover, the mutation was detected in the mRNA (Figure 10D), further indicating that P186L compromised ZCCHC8 protein levels. To test this directly, we quantified the stability of transfected Myc-tagged wild-type and mutant ZCCHC8 and found p.P186L mutant levels were lower, supporting the conclusion that this mutation disrupts protein stability (Figure 10E). We next tested whether ZCCHC8 loss affects TR levels. We generated HeLa cells that stably express an shRNA against ZCCHC8 and found that TR levels were significantly reduced, similar to NAF1 knockdown (Figure 10F-H). The decrease in total TR levels was detected by qRT-PCR as well as northern blot (Figure 10G-H and TR northern specificity validated and not shown). These data indicated that *ZCCHC8*^{P186L} is a loss-of-function mutation and that ZCCHC8 loss is sufficient to cause TR insufficiency.

Figure 10. Pulmonary fibrosis-associated mutation affects ZCCHC8 stability. A. Immunoblot of endogenous ZCCHC8 levels in lymphoblastoid cell lines (LCLs) from healthy controls (C1 and C2), unaffected relatives and mutation carriers as indicated. Result replicated in two other blots. Pedigree identifiers refer to Figure 1A. B. Immunoblot showing ZCCHC8, SKIV2L2 and RBM7 levels in primary



skin fibroblasts from the proband. C. Total *ZCCHC8* mRNA levels in LCLs from unaffected relatives (n=4) and *ZCCHC8* p.P186L mutation carriers (n=3). D. Chromatogram showing *ZCCHC8* p.P186L mutation is detectable in mRNA from LCLs. E. Immunoblot of transfected Myc-tagged (into 293FT cells) and endogenous ZCCHC8 showing the P186L mutation is less stable. F. Immunoblot showing efficacy of shRNA knockdown of Luciferase (Luc), ZCCHC8 and NAF1 in HeLa cells on protein levels. G. Total *TR* levels measured by qRT-PCR after knockdown. Mean is from 3 independent knockdowns. H. Northern blot shows low *TR* levels after stable knockdown of *ZCCHC8* and *NAF1*. Quantification is shown below each of the gels and each experiment was replicated at least twice where not noted. For C and G, the mean ± SEM is shown. **P<0.01 and ***P<0.001 (Student's *t*-test, two-sided).

ZCCHC8 is required for nascent TR 3'-end processing and for telomerase function

To test the consequences of ZCCHC8 loss on TR integrity and telomerase function, we generated *ZCCHC8* null HCT116 pseudodiploid cells by CRISPR/Cas9 editing (Figure 11A-B). The NEXT complex is thought to target a subset of nascent RNAs for 3'end processing by the nuclear RNA exosome (*37*) and we hypothesized that TR's 3'ends may be misprocessed in mutant cells. We used a next generation sequencing approach that characterized TR's 3'ends by Rapid Amplification of cDNA Ends followed by sequencing (3'RACE-seq) (Figure 11C). We found that compared to isogenic controls, *ZCCHC8*$^{-/-}$ cells had decreased mature TR fraction (451 nt long) and increase in 3'extended TRs (Figure 11D). The longer forms included adenylated mature TR as well as extended forms which mapped to the genomic locus beyond the annotated mature TR 3'-end, consistent with their being nascent precursors (Figure 11D). We categorized these extensions based on the bioinformatic tools we used as short (<15 nt beyond the mature 3'end) and long (>15 nt), and we found *ZCCHC8*$^{-/-}$ cells had increased levels in both of these categories. We directly quantified the longer TR forms by qRT-PCR and found 4- to 10-fold higher levels of these TRs which extended +20, +51, and as far as +784 nucleotides beyond the mature TR 3'end (Figure 11E). Total TR levels were also low in *ZCCHC8*$^{-/-}$ cells by northern blot (mean 79%, P=0.038, Student's t-test, Figure 11F-G. TR levels measured by northern blot may however over-estimate the fraction of functional TR since a significant of TRs in *ZCCHC8*$^{-/-}$ cells were 3'extended. We therefore directly measured telomerase enzyme activity using the telomere repeat amplification protocol (TRAP) assay, and found that, similar to NAF1 mutant cells (*18*), *ZCCHC8*$^{-/-}$ cells had decreased TRAP activity (mean 52% of isogenic ZCCHC8+/+ cells, P=0.001, Student's t-test, Figure 11H-I). These data showed that ZCCHC8 is required for TR 3'end processing and for telomerase enzyme function.

To test the clinical relevance of these observations, we performed 3'RACE-seq on the proband's primary fibroblasts and found they had a similar TR length distribution as *ZCCHC8*$^{-/-}$ cell lines (Figure 11J). There was a relative decrease in mature TR and, conversely, an increase in the TR fraction with short (<15 nt) and long (>15 nt) extensions beyond the mature 3'end (Figure 11J). Quantification by RT-PCR also confirmed these extensions at +20, +51 and +784 nucleotides beyond the mature TR 3'end (Figure 11K). We, however, found heterogeneity in the stability of these longer TR forms across cell types since we could not amplify them readily from Epstein-Barr virus-transformed lymphoblastoid cells, consistent with data suggesting herpes viruses can interfere with RNA turnover in hematopoietic-derived cells (*73*).

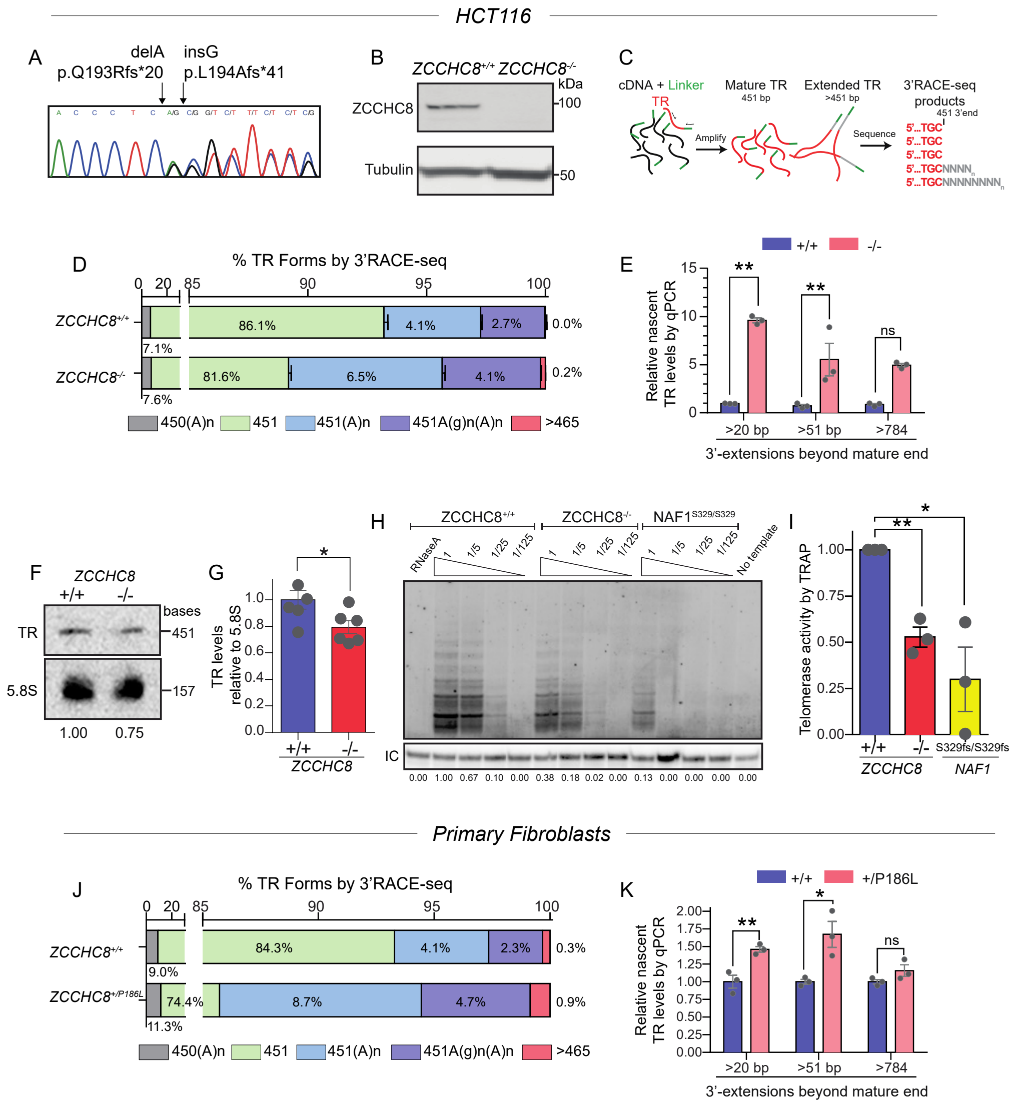


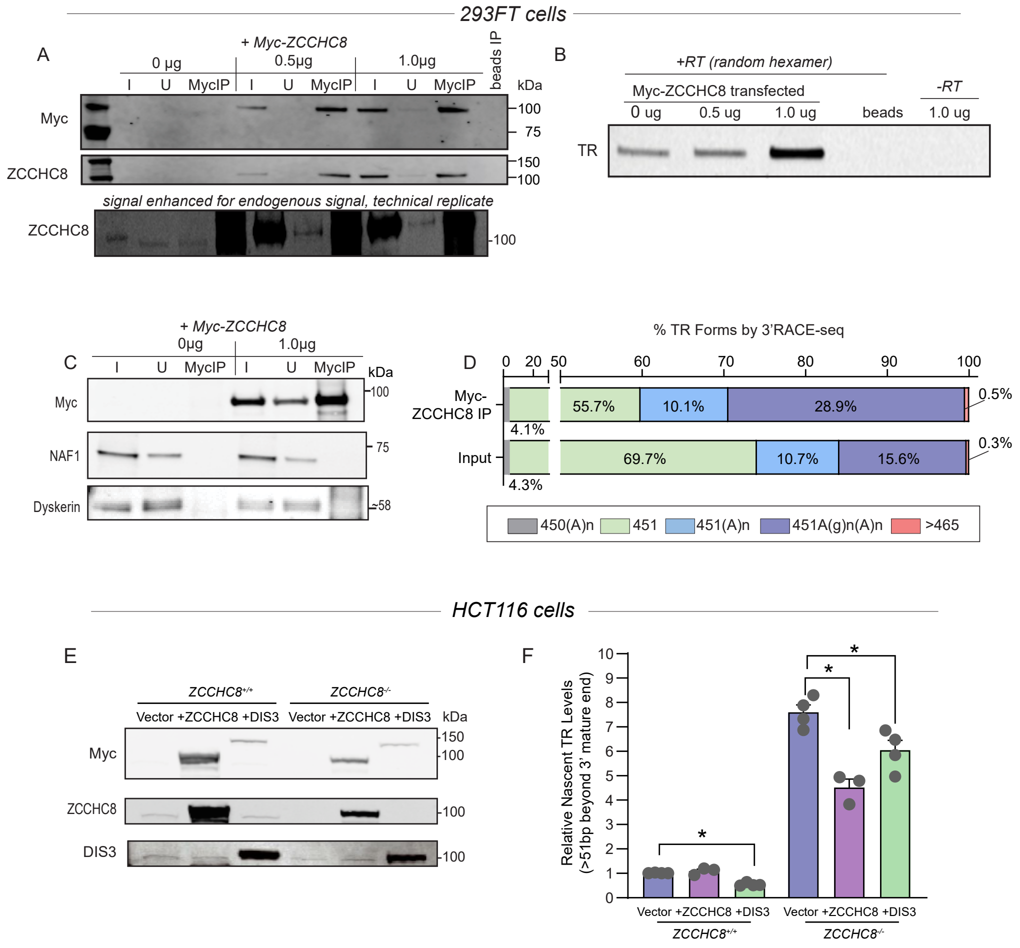
Figure 11. Pulmonary fibrosis-associated mutation affects ZCCHC8 stability. A. Chromatogram showing compound heterozygous frameshift (fs) mutations generated using CRISPR/Cas9 in HCT116 pseudodiploid cell lines. B. Immunoblot for ZCCHC8 in HCT116-edited cells. C. Schema summarizing *TR* 3' Rapid Amplification of cDNA Ends sequencing (3'RACE-seq) method. *TR* 3'ends were generally divided into mature (451 bp) and extended (>451 bp) where extensions are denoted by gray N's. D. Summary of *TR* 3'RACE-seq fractions in *ZCCHC8*$^{+/+}$ and *ZCCHC8*$^{-/-}$ HCT116 cells. The color-coded key shows the different categories of *TR* forms including adenylated and genomically-extended forms that are adenylated and long extended forms that are >465 nucleotides. Data reflect mean of 3 independent 3'RACE-seq analyses each from a different RNA isolation. *ZCCHC8*$^{+/+}$ cells were isogenic having undergone selection. E. qRT-PCR of extended *TR* forms beyond the 451 mature end (>20, >51, >784 nucleotides). Data reflect mean values from three independent RNA isolations. F. Northern blot for *TR* of edited *ZCCHC8*$^{+/+}$ and *ZCCHC8*$^{-/-}$ HCT116 cells. G. Quantification of six northern blots from three

independent RNA isolations. H. Telomerase activity measured by Telomere Repeat Amplification Protocol (TRAP) assay in *ZCCHC8*$^{+/+}$, *ZCCHC8*$^{-/-}$, and *NAF1*$^{S329/S329}$ HCT116 cell extracts. Activity was quantified in serially diluted extracts (1, 1/5, 1/25, 1/125) against a PCR-amplified internal control (IC). An RNAse treated *wild-type* extract and no template PCR reaction are negative controls. I. Mean TRAP activity of 1/5x diluted extracts from 3 independent TRAP assays each from a different cell lysate. J. Summary of 3'RACE-seq of *TR* forms from control and proband primary skin fibroblasts as in (D). K. Mean qRT-PCR values of extended *TR* forms in primary skin fibroblasts as in (E) (n=3 technical replicates). Data are expressed as means ±SEM. $^{*}P < 0.05$ and $^{**}P < 0.01$ (Student's *t*- test, two-sided).

ZCCHC8 associates with TR precursors and nascent TR accumulation is rescued by a catalytic RNA exosome component

Our genetic data indicated that ZCCHC8 is required for 3'end processing of TR. To test if ZCCHC8 associates with TR, we performed RNA immunoprecipitation (IP) with Myc-tagged ZCCHC8 in 293FT cells (Figure 12A) and found TR was enriched in the immunoprecipitated fraction (Figure 12B). We considered whether this amplification may be secondary to an association with the dyskerin complex, but we could not co-immunoprecipitate dyskerin or NAF1 with Myc-ZCCHC8 (Figure 12C). We then examined whether ZCCHC8 interacts with TR precursors by performing TR-specific 3'RACE-seq of ZCCHC8-immunoprecipitated RNA. We found an enrichment of longer TR forms containing genomically-encoded 3'ends in the ZCCHC8-bound RNA IP fraction (Figure 12D). The majority of the extensions were shorter than 15 nt (Figure 10D). We directly tested whether the accumulation of TR precursors may be rescued by forced expression of DIS3, a 3'>5' exoribonuclease of the RNA exosome (Figure 12E). Exogenous expression of ZCCHC8 rescued the accumulation of nascent TR in HCT116 *ZCCHC8*$^{-/-}$ cells, as expected for an on-target effect. Over-expression of DIS3 also had a similar effect in reducing these extended TR forms (Figure 12F). These data supported that nascent TR accumulation in ZCCHC8 deficient cells is linked to inadequate nuclear RNA exosome targeting likely through DIS3.

Figure 12. ZCCHC8 immunoprecipitates *TR* precursors with genomically encoded 3'ends. A.



Immunoblot showing input (I) and immunoprecipitated (IP) Myc-tagged ZCCHC8 plasmid transfected at two concentrations into 293FT and harvested at 48h. The lower most panel is an over-exposed blot to show the relative levels of tagged ZCCHC8 relative to endogenous. B. Amplified *TR* from I and IP fractions using primers within the mature *TR* sequence. C. Western blot of IP Myc-tagged ZCCHC8 shows no detectable dyskerin or NAF1 co-IP. D. Deconvoluted *TR* 3'RACE-seq data showing enrichment of genomically encoded *TR* precursors in the Myc-ZCCHC8 fraction compared to input. The genomic TR locus with the 451 mature 3'-end is annotated in the left upper corner. In the color-coded key of the *TR* species, Misc refers to non-genomic, non-templated nucleotides other than 'A.' E. Western blot showing transfection of Myc-tagged *ZCCHC8* and *DIS3* plasmids into CRISPR/Cas9 edited HCT116 cells (Figure 3A). E. qRT-PCR of nascent *TR* that is >51 nucleotides extended beyond the 3' mature *TR* end with transfected tagged ZCCHC8 and DIS3.

DISCUSSION

Using linkage analysis in a family with autosomal dominant pulmonary fibrosis, we identified a novel regulator of telomere maintenance and *TR* 3'end processing. We show ZCCHC8 is required for *TR* function by mediating its 3'end maturation through targeting to the nuclear RNA exosome. Our data are notable, since we show that, while ZCCHC8 is required for the processing of other RNAs, heterozygous loss is tolerated and, in humans, selectively manifests as a short telomere phenotype including pulmonary fibrosis. In yeast, the most common functional category of telomere length regulators is RNA processing genes which account for nearly one-third of all hits (*12-14*). Our data support the short telomere phenotype, especially familial pulmonary fibrosis, being enriched for mutant regulators of *TR* integrity.

The NEXT complex nor the RNA exosome has not been previously implicated in telomerase RNA maturation. Nascent telomerase RNA has been shown to accumulate upon knockdown of various core RNA exosome components and RBM7 but only been implicated as a step in degradation. Further those studies were limited to aneuploid cancer cell lines where RNA turnover and nascent telomerase RNA accumulation may vary from those of primary human cells and mask an underlying defect in a very low abundant RNA (*30, 40*). Specifically ZCCHC8 has not been shown to affect telomerase RNA processing either in maturation or degradation in transient siRNA knockdown HeLa studies (*40*). Here we show upon stable knockdown of telomerase RNA in multiple independent transduction experiments across multiple assays, both qRT-PCR and northern blot analysis, total TR is significantly decreased. Further, nascent transcripts accumulate but regardless, telomerase activity is compromised despite these accumulations. This is similar to findings when loss or mutation of the polyadenylase PARN results in accumulation of nascent TR with compromised telomerase homeostasis in patients and cell lines (*17, 30*). Given this and our findings, we would hypothesize 3’ extended transcripts to be nonfunctional. Future **in vitro direct telomerase assay** with extended transcripts of different lengths can definitively answer this question and be an important step in solidifying the functional characteristics and processing of nascent TR.

The implication of the NEXT complex and RNA exosome expands our understanding of TR 3’-endprocessing. Prior to these discoveries presented here defective TR 3’ end processing in *PARN* mutant patient cells resulting in accumulation of short, polyadenylated TR transcripts caused a short telomere phenotype in heterozygous and biallelic families (*16, 17*). These findings implicated deadenylation into the maturation pathway of TR; however, the question of how nascent TR is trimmed to the most common “mature” 451 bp form has remained elusive. Longer TR forms have been reported and detected in cell lines (*30*)but how or whether these longer nascent TR forms are trimmed to the shorter mature forms was unclear. Here we propose a model by which nascent TR is transcribed by RNA polymerase II to extended TR forms that are subsequently trimmed and processed back at the 3’end to functional 451 bp form. Nascent TR forms are likely polyadenylated as are other RNA pol II transcripts and targeted either to maturation or degradation. We propose that ZCCHC8 via the NEXT complex targets long, genomically extended forms towards maturation. This likely involves additional processing by *PARN* to remove polyadenylated tails. Failure to do so results in accumulation of both genomically extended and polyadenyalted nascent TR forms that are subsequently targeted to degradation. Without proper trimming and 3'-end processing by ZCCHC8 and the NEXT complex, mature TR forms are not produced and TR is unable to serve as a template for telomere elongation in its immature forms.

METHODS

Subjects and study approval. Patients were recruited from 2007-2017 as part of the Johns Hopkins Telomere Syndrome Registry as previously described (*74*). The study was approved by the Johns Hopkins Medicine Institutional Review Board and all the subjects gave written informed consent. The proband was identified in a screen for individuals who had low TR levels in lymphoblastoid cell lines (LCLs) by qRT-PCR as previously described (*18*). LCLs were generated by Epstein Barr virus transduction (*75*), and primary skin fibroblasts from the index case were derived from a skin punch biopsy using standard methods. Control skin fibroblasts were purchased from ATCC (BJ, CRL-2522). The screen of 42 patients for mutant *ZCCHC8* included genetically uncharacterized familial pulmonary fibrosis-emphysema probands (n=31) and patients with a classic short telomere syndrome phenotype and short telomere length as previously defined (n=11) (*18*).

Telomere length measurement and DNA extraction. Telomere length was measured on peripheral blood-derived cells by flow cytometry and fluorescence in situ hybridization (flowFISH) at the Johns Hopkins Pathology Laboratories as described (*74, 76*). Genomic DNA was extracted from fresh blood or PBMCs using the Gentra Puregene Kit B (Qiagen) and from formalin-fixed paraffin embedded (FFPE) using the FFPE Tissue DNA Isolation Kit (MO BIO Laboratories).

Quantitative Real Time-PCR (qRT-PCR) for TR. RNA levels were measured using the SYBR Green qRT-PCR method as previously described (*18*). Briefly, RNA was isolated (RNeasy, Qiagen) and reverse transcribed using random hexamer primers and Superscript III (Invitrogen). Within any given experiment, RNA isolation and cDNA reverse transcription were side-by-side for all the samples. TR levels were normalized to *ARF3* (human). Primers for quantifying *hTR* extensions were from (*30, 40*).

Whole genome sequencing. Whole genome sequencing was performed on blood-derived DNA (Macrogen) using a HiSeq X platform (Illumina). Variants were called and aligned to GRCh37 reference genome using the Isaac variant caller (v.2.0.13) and aligner (v.01.15.02.08). The mean depth of coverage was 48X with 98% of the genome covered at 20X and 93% at 30X. Annotation and filtering was performed using PhenoDB (*77*). Non-synonymous exonic and splice junctions were defined as rare if they were absent in dbSNP builds 129, 131 and 135, and also had minor allele frequency (MAF) less than 0.0001 in 1000 genomes and the NHLBI GO Exome Sequencing Project (accessed June 19, 2017 and updated January 6, 2019).

Genome wide linkage analysis. To identify the disease gene, we genotyped a total of 14 individuals using a SNP array (548K, Infinium HumanCoreExome-24v1-0, Johns Hopkins Genetic Center Resource Facility). SNPs were ranked in 400 kb bins by quality metrics based on call rates, position and tightness of the genotype cluster. The 6,710 high quality autosomal SNPs were then used in a singlepoint linkage analysis using MERLIN software (*78*). Subjects were assigned 'affected' status if they had a short telomere syndrome phenotype such as IPF, were obligate carriers, or had low *TR* levels and abnormally short telomere length. We used an autosomal dominant model with prevalence of 0.0001, and penetrance of 1.0 and 0.0001 in carriers and non-carriers, respectively.

ZCCHC8 multiple species alignment. Human ZCCHC8 (NP_060082) was aligned to the following vertebrate ZCCHC8 protein sequences (*Canus familiaris*: XP_005636220; *Mus musculus*: NP_081770; *Bos taurus*: NP_001192619; *Gallus gallus*: NP_001192619, *Xenopus tropicalis*: NP_001135573; *Oryzias latipes*: XP_004072626; and *Danio rerio*: NP_001077287) within Jalview (version 2.10.2b2) using Clustal Omega (*79*).

Targeted sequencing of ZCCHC8. The heterozygous *ZCCHC8* P186L variant was verified using Sanger sequencing of the 102 bp magnetic bead purified PCR product (AMPure XP, Beckman Coulter Life Sciences) using primers available upon request. To screen for *ZCCHC8* mutations in short telomere syndrome patients, we sequenced gDNA for the coding sequence and exon-intron junctions on a customized targeted nextgen sequencing panel (Truseq, Illumina) as previously (n=21 cases) (*42*). The mean depth of coverage was 50X. For remaining cases (n=21), we Sanger sequenced the *ZCCHC8* cDNA derived from LCLs bi-directionally (primers available upon request).

Western blots. Total protein was isolated from cells in RIPA buffer (Cell Signaling Technology) with cOmplete Mini protease inhibitor (Sigma-Aldrich) and lysates were quantified using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Using the NuPAGE SDS-PAGE gel system, protein (10-20µg) was run on 10% Bis-Tris gels with MOPS-SDS running buffer at 150V and transferred to a PVDF membrane by wet transfer using XCell SureLock Mini-Cell Elecrophoresis System or by the iBlot 2 Dry Blotting System at 30V (Thermo Fisher Scientific). Immunoblotting was performed using the LI-COR system as previously (*18*). Membranes were blocked for 1h to overnight prior to antibody staining. The following antibodies were used with clone identifiers listed for the two monoclonal antibodies (Actin and Myc): Actin (mouse, ab8226, 1:2000; Abcam, mAbcam8226), DIS3 (rabbit, ab179933, GR:195524-7, 1:500, Abcam), Myc (mouse, clone 4A6, 1:1000; Millipore, 4A6), NAF1 (rabbit, ab157106, 1:1000, Abcam), RBM7 (rabbit, HPA013993, 1:250, Sigma), SKIV2L2 (rabbit, ab187884, 1:500, Abcam), Tubulin (rabbit, ab6046, 1:5000; Abcam), ZCCHC8 (mouse, ab68739, 1:500, Abcam). Membranes were stained with anti-mouse or anti-rabbit IRDye secondary antibodies (IR680 or IR800, donkey, 1:10,000 for both) before visualization on an Odyssey scanner (LI-COR). Quantification of protein levels was done using ImageJ (*80*).

Northern quantification of TR. We optimized a protocol for quantifying TR levels using a modified northern protocol to maximize visualization of TR, a low abundance RNA. Total RNA (0.75-6µg) was mixed with RNA loading buffer (Sigma R4268-1VI) and heated to 95°C for 3m then cooled on ice for 3m. Samples were then immediately loaded onto a pre-run Novex 6% TBE-UREA gel and run at 180V in heated TBE buffer (65°C) for 55m on an Isotemp hot plate set to 70°C (Thermo Fisher Scientific). Gels were transferred at 20V onto an Amerhsam Hybond-XL membrane for 90m (GE Healthcare Life Sciences). The cross-linking, hybridization and probe preparation were performed as previously described for *hTR* and *5.8S* (*18*).

Cloning and site directed mutagenesis. Human *ZCCHC8* (NM_017612), *RBM7* (NM_001286045), *SKIV2L2* (NM_015360), and *DIS3* (NM NM_014953.4) were cloned from total cDNA with addition of an N-terminal Myc-DDK tag into a CMV promoter driven pcDNA5/FRT/TO expression vector using restriction digestion and Gibson cloning (New England Biolabs). Plasmids were transfected using Lipofectamine 2000 or 3000 (Invitrogen) and protein lysates were harvested after 24-48h.

Transduction, lentivirus production and shRNA knockdown. Virus production in HEK-293FT cells, as well as the stable lentiviral shRNA knockdown in HeLa cells, were performed as previously described (*18*). shRNAs were in PLKO.1 cloning vectors: Luciferase (SHC007, Sigma), NAF1 (TRCN000135721, Open Biosystems) and ZCCHC8 (TRCN0000075158, Open Biosystems).

Generation of ZCCHC8 null human cells. We used CRISPR/Cas9 editing to generate *ZCCHC8* knockout HCT116 cells. To first generate a *ZCCHC8* knockout line, the sgRNA (5’ cccatccttcggaaagctga 3’, 20 bp) was cloned into the pX458 vector (gift from Feng Zhang, Addgene #62988) (*81*) and transfected into pseudodiploid HCT116 cells. GFP-positive cells were flow-sorted in bulk and re-plated using single-cell dilutions. Mutants were screened for by PCR and Sanger sequencing (F: 5' acacatcaaagtctggctcctt 3', r: 5' cggaaagctgagggttttc 3').

Cell line authentication. HCT116 cells (gift from Dr. Ben Ho Park, Johns Hopkins University) were authenticated by short tandem repeat profiling according ATCC guidelines using the PowerPlex 16HS short tandem repeat profiling kit (Promega) and matched ATCC# CCL-247 HCT116 human colon carcinoma lines according to the ANSI/ATCC ASN-0002-2011 standard (ANSI eStandard) (Johns Hopkins Genetic Resources CORE Facility).

3'Rapid Amplification of cDNA Ends sequencing (3'RACE-seq). 3'RACE for *TR* was performed as previously described (*17, 29*) with modifications. Briefly, DNase I-treated (Qiagen) total RNA was ligated to 5μM pre-adenylated linker (Universal miRNA Cloning Linker, New England Bio Labs)with 280U T4 RNA ligase 2, truncated KO (New England BioLabs) in a 20μl reaction at 25°C for 16h with RNaseOUT and 25% PEG8000 (New England BioLabs). The ligation reaction was then cleaned using RNA Clean and Concentrator (Zymo Research) and cDNA was synthesized with a universal RT primer (5'-CTACGTAACGATTGATGGTGCCTACAG) using the SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). *TR* PCR was carried out using hTR_RACE_F and hTR_RACELinker_R (final 0.4 μM) (primer sequences available upon request), dNTPs (final 200 μM) and Q5 high fidelity DNA Polymerase (at 0.02 U/μl) (New England Biolabs) with these parameters: 98°C for 1m, 22 cycles of 98°C for 10s, 67°C for 30s, and a final extension at 72°C for 2 min. The product was then purified (QIAquick, Qiagen) and visualized on agarose gels for quality control prior to library preparation. Libraries were prepared using the TruSeq Nano DNA LT Library Prep kit (Illumina) following the manufacturer's instructions but starting with adenylation of 3'ends. Samples were quantified by Bioanalyzer diluted to 2nM each then pooled. The combined libraries, containing 15-20% phiX control (Illumina), were then run on an Illumina MiSeq with 250 paired-end reads (Johns Hopkins Genetic Resources Core Facilities).

3'RACE-seq analysis. Reads were demultiplexed and the read 1 and read 2 fastq files for each sample were generated. Reads were trimmed using a custom script. To quantify the length and sequence of short extensions (defined as up to 15 bases beyond the 451st nucleotide of the mature hTR end with or without an oligoA tail), reads were mapped to the reference genomic *TR* locus (NR_001566) appended with the linker sequence using Bowtie2(*82*) and extensions were determined as insertion blocks in the alignments. The mean number of reads mapped per sample was 200,745 (range, 79,000-301,000) and the median alignment rate was 80% (range, 42-89). Because of the limited amount of RNA, the 3'RACE-seq from the ZCCHC8 RNA IP had a lower alignment (17%), corresponding to 132,550 reads. Using a customized script, reads were then filtered to contain both at least 11 bases of the 3' end of mature *TR and* at least 9 of the 17 bases of the linker sequence. The relative abundance of different 3' termini for *TR* was determined by normalizing to the percentage of total trimmed and filtered reads.

Long genomic extensions (i.e. >15 bases beyond the mature *TR* end) were quantified separately using Bowtie and Sim4db (*83*) to ensure mapping to the entire *TR* locus including +/- 5 kb sequences flanking the mature *TR*. SAMtools were subsequently used to extract reads corresponding to the extended region >15 bases beyond the end of mature *TR*. The mean number of reads mapped was 181,460 (range, 70,000-282,000) and the median alignment rate was 71% (range, 31-85). Percent alignment for RNA immunoprecipitate (IP) was 16% with 124,020 reads aligning to the target *TR* sequence. Customized scripts for *TR* 3'RACE-seq analysis are available upon request.

Telomerase Repeat Amplification Protocol (TRAP). TRAP was performed as previously described (*84*). HCT116 cells were treated with CHAPS lysis buffer (10^4 cells/μL) and isolated protein was quantified using a BCA assay as above.

Protein and RNA IP. 293FT cells were grown to 75-90% confluence and transiently transfected with pcDNA5/FRT/TO/ZCCHC8-Myc-DDK plasmid using lipofectamine 2000 reagent (ThermoFisher). After 48h, cells were washed with PBS supplemented with protease inhibitors (Roche) and lysed cells in buffer [20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP40, 2mM MgCl2, 1x EDTA-free Halt Protease, phosphatase inhibitor cocktail (ThermoFisher) and RNAsin (Promega)]. The lysate was rotated end-to-end at 4°C for 15m and clarified by spinning at 4°C for 10m at 12.3 rpm. The supernatant lysate was incubated with anti-c-Myc magnetic beads (Pierce) for 2h at room temperature, rotating end-to-end. After washes in 1x TBS-T (25mM Tris, 0.15M NaCl, 0.05% Tween-20) and lysis buffer, one third of the sample was processed for protein IP and the rest for RNA IP. For protein IP, the bound protein was eluted with NuPAGE LDS Sample Buffer (4X) and then diluted to 2X with Lysis Buffer. For RNA IP, the beads were resuspended in TriZol, incubated at -80°C overnight and RNA was extracted as a co-precipitate based on the manufacturer's protocol (Life Technologies). The isolated RNA was treated with DNAse (Ambion, AM1906) and used as a template for cDNA synthesis using Superscript III reverse transcriptase (Invitrogen) and random hexamer primer mix (Thermo Fisher Scientific). *TR* was then amplified by RT-PCR using the same primer set used for total *TR* quantification described elsewhere here. Alternatively total RNA was used for *TR* 3'RACE-seq.

Statistical analyses. Singlepoint parametric linkage was analyzed using logarithm of the odds scores. qRT-PCR analyses were reported as standard error of the mean (s.e.m.), and *P* values were calculated using GraphPad Prism software. For each PCR run, samples were done in triplicate. All P-values shown are two-sided. Statistical tests uses are indicated in the text and legends with the respective data.

Chapter Four

Loss of ZCCHC8 in mice leads to severe neurodevelopmental defects

Dustin L. Gable[1,2,3], Valeriya Gaysinskaya[2,3], Christine C. Atik[2,3], C. Conover Talbot Jr.[4], Byunghak Kang[5], Nuria Amat-Codina[2,3], Cory Brayton[5], Liliana Florea[6], Mary Armanios[2,3,6,7]

Medical Scientist Training Program[1]
Department of Oncology[2]
Telomere Center[3]
Institute for Basic Biomedical Sciences[4]
Department of Comparative and Molecular Pathobiology[5]
McKusick-Nathans Institute of Genetic Medicine[6]
Sidney Kimmel Comprehensive Cancer Center[7]
Johns Hopkins University School of Medicine
Baltimore, MD 21287

INTRODUCTION

We have shown that ZCCHC8 knockdown in HeLa cancer cell lines is sufficient to decrease TR levels and HCT116 cancer cells null for ZCCHC8 are viable but deficient for telomerase activity; however, additional RNA processing defects secondary to complete ZCCHC8 loss remain unknown. Previous work in HeLa cells elucidating additional ZCCHC8 functions are limited but support a role for ZCCHC8 in non-nucleolar localization and processing of PROMoter uPstream Transcripts (PROMPTs). Other RNA targets of ZCCHC8 and consequences of its loss in connection with the NEXT complex have remained elusive requiring further study (*37*). Our studies in NAF1 mutant mice highlight the dosage sensitivity of telomerase RNA (TR) stability such that heterozygous loss selectively disrupts telomerase RNA levels with sparing of RNA pseudouridylation while homozygous loss is embryonic lethal pointing to the essential NAF1 functions beyond telomerase RNA maintenance likely in rRNA modifications given TR null mice are viable (*18*). Similarly we would hypothesize that additional regulators of telomerase RNA biogenesis including ZCCHC8 would exhibit TR dosage sensitivities such that heterozygous loss perturbs telomerase RNA homeostasis sufficient to cause telomere shortening whereas homozygous loss has additional biological consequences beyond defective processing of nascent TR.

In order to perform hypothesis driven experiments and identify consequential RNA processing defects secondary to ZCCHC8 loss in a biologically and clinically relevant model we engineered ZCCHC8 null mice in a C57BL/6J background. Observed phenotypes would then be secondary to non-telomeric function of ZCCHC8 given the long telomeres (40 kb) in this strain of mice that would require substantial number of backcrosses for any relevant measurable phenotype effect seen in mice with telomere lengths similar to human (*55, 85*). In both mice and humans, TR is haploinsufficient for telomere maintenance. TR+/- mice with telomere lengths similar to humans show genetic anticipation, and late generation mice with short telomeres develop bone marrow failure and other short telomere syndrome phenotypes (*6, 15, 55*). While Zcchc8+/- mice had TR insufficiency but were otherwise intact, as expected for first generation telomerase null mice, Zcchc8-/- mice showed severe neurodevelopmental defects that overlapped with features of a ciliopathy. In these Zcchc8-/- mice, there was a pervasive RNA dysregulation defect that included accumulation as well as mis-processing of RNAs encoding cilia components as well as replication dependent histones (RDH). These transcripts shared a similar length to TR. Our data indicate that vertebrate TR requires ZCCHC8 for its maturation, along with other low abundance RNAs. In contrast to complete ZCCHC8 loss, however, which manifests as a pervasive transcriptome defect, heterozygous mutations preferentially affects TR and manifests in TR insufficiency and a familial short telomere syndrome.

RESULTS

Zcchc8 **loss causes a gene dose-dependent misprocessing and overall insufficiency of** ***TR***

To test the consequences of ZCCHC8 loss in vivo, we used CRISPR/Cas9 to edit the mouse *Zcchc8* locus (Figure 13A). We were particularly interested in this question since we noted mention in the literature of one family with homozygous null mutation in *ZCCHC8,* L90X, that was identified in a study examining genetic determinants of autosomal recessive intellectual disability in consanguineous families (*86*). The genome editing strategy we designed introduced a null allele with a premature stop codon at amino acid 93 in exon 2 of mouse *Zcchc8* (Figure 13A). We interbred heterozygous mutant mice and detected viable heterozygous and homozygous null progeny that had half and no detectable ZCCHC8, respectively (Figure 13B). These data suggested that *Zcchc8* is not essential for embryogenesis (Figure 13B). In cells derived from adult *Zcchc8*$^{+/-}$, we found the levels of other NEXT complex proteins, RBM7 and SKIV2L, were preserved similar to the human patient (Figure 13B), while *Zcchc8*$^{-/-}$ knockout mice had minimally decreased levels of these proteins (Figure 13C and antibody cross-reactivity validated and data not shown). These data suggested that the stability of RBM7 and SKIV2L2 does not depend on ZCCHC8 integrity. To assess the consequences of *ZCCHC8* loss on *TR*, we measured total *TR* levels by northern blot and found that, similar to *mTR*$^{+/-}$ mice, total *TR* was decreased in both *Zcchc8*$^{+/-}$ and *Zcchc8*$^{-/-}$ mice (P=0.02 for both, Student's *t*-test, Figure 13D-E). Because *TR* in these *Zcchc8* mutant mice may include aberrant non-functional forms, total levels by this northern analysis may under-estimate the extent of the *TR* insufficiency in these mice. We specifically quantified *mTR* precursors by amplifying 3'extended products that were at least 20 nt long by qRT-PCR and found that *Zcchc8*$^{+/-}$ and *Zcchc8*$^{-/-}$ had a dose-dependent increase with *Zcchc8*$^{-/-}$ mice having the highest levels (Figure 13F). These data affirmed that ZCCHC8 plays a role in vertebrate *TR* 3'end processing and that this effect is dose-dependent.

Zcchc8$^{-/-}$ mice develop a progressive fatal pathology

We next examined the vivo consequences of *Zcchc8* loss. The interbreeding of *Zcchc8*$^{+/-}$ mice revealed a distortion of Mendelian ratios in their progeny with *Zcchc8*$^{-/-}$ representing only 5% of adult mice surviving to day 30 ($P<0.001$, chi-square test, $n=170$ mice genotyped, Figure 13G). The viable mice developed a progressive decline and none survived beyond 70 days. Their attrition occurred postnatally as *Zcchc8*$^{-/-}$ Mendelian ratios were preserved for both E12.5 embryos as well as newborn pups (Figure 13G).

***Zcchc8*$^{+/-}$ mice are intact but *Zcchc8*$^{-/-}$ littermates exhibit a progressive neurodevelopmental defect**

To characterize the consequences of ZCCHC8 loss in vivo, we first studied adult *Zcchc8*$^{+/-}$ mice and found they were grossly and histologically indistinguishable from *wild-type* littermates (Figure 14A). Specifically, we found no abnormalities in the organs examined including the brain, viscera and hematopoietic system (n=9/genotype examined, 5 males/4 females). These observations were consistent with *Zcchc8*$^{+/-}$ mice primarily having *TR* insufficiency that is predicted to cause telomere shortening after many successive generations as has been documented for *mTR*$^{+/-}$ and *mTR*$^{-/-}$ mice (*6, 87*).

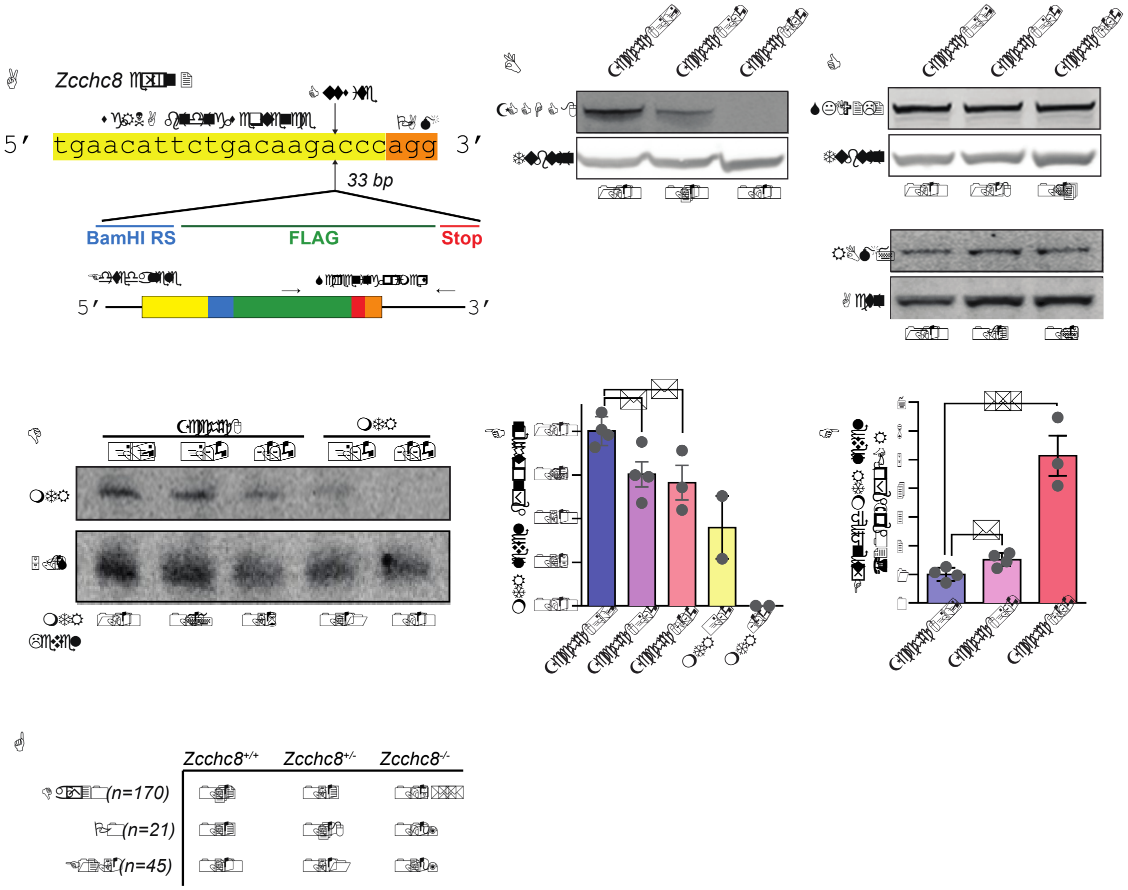


Figure 13. Design of *Zcchc8* targeting strategy using CRISPR/Cas9. A. Scheme depicts CRISPR targeting of exon 2 within the sgRNA binding site that is 3 base pairs upstream of the PAM sequence to incorporate a 33 base pair insert (arrows) containing a stop codon (red). *BamH1* restriction site and the FLAG sequence identify edited alleles. The screening strategy uses a forward primer in the edited exon and a downstream intronic reverse primer (black arrows). B and C. Immunoblot for ZCCHC8 and SKIV2L2 and RBM7 protein in mouse skin fibroblasts. D and E. Northern blot for mouse *TR* (*mTR*) levels quantified relative to 5.8S rRNA in skin fibroblasts. For (E), mean reflects mice *Zcchc8*$^{+/+}$ (n=4, 2M/2F), *Zcchc8*$^{+/-}$ (n=4, 2M/2F), *Zcchc8*$^{-/-}$ (n=3M), *mTR*$^{+/-}$ (n=2, sex unknown) and *mTR*$^{-/-}$ (n=2, sex unknown). F. *mTR* 3'-extended levels (>20 bp) relative to Hprt as measured by qRT-PCR. Mouse numbers and sex designations are as in E. Data in B-F are from mouse adult ear fibroblasts. G. Genotype frequency from *Zcchc8*$^{+/-}$ interbreeding of E12.5 embryos, newborn pups (P0), and day 30 adult mice with the number of each genotype screened shown. Data are expressed as means ±SEM. *P < 0.05, **P < 0.01, and ***P < 0.0001 (Student's *t*- test, two-sided).

In contrast, adult *Zcchc8*$^{-/-}$ mice had evident cranial abnormalities with domed-shaped heads, a sign of hydrocephalus (Figure 14A, top panel). Computed tomography (CT) imaging confirmed cranial deformities with dilated cranial sutures (Figure 14A, middle and lower panels, respectively). We performed a detailed histologic analysis and found that nearly all *Zcchc8*$^{-/-}$ mice that survived at least 30 days had severe hydrocephalus with ventriculomegaly (9 of 11, Figure 14B). In mice, hydrocephalus is a common manifestation of ciliopathy; it arises because of defects in ependymal motile cilia that in turn impair cerebrospinal fluid movement (*88*). We therefore examined *Zcchc8*$^{-/-}$ mice for other ciliopathy features and found a high prevalence of severe otitis media, another phenotype caused by defects in motile cilia lining the middle ear mucosa (8 of 10). We, however, did not find evidence for laterality defects, such as situs inversus or heterotaxy (data not shown). Moreover, there was no evidence of upper airway disease as is seen with some motile ciliopathies. To assess the cause of the cranial defects, we traced brain development in *Zcchc8*$^{-/-}$ embryos and found they had grossly small crania and brain volumes, indicating a defect in neurogenesis preceding the onset of hydrocephalus post-natally (Figure 14C-D). *Zcchc8*$^{-/-}$ embryos and newborn pups were also microcephalic, but developed macrocephaly by the time they reached adulthood (Figure 14F-G). These data indicated that *Zcchc8*$^{-/-}$ mice develop a progressive neurodevelopmental pathology which shares some features with ciliopathies.

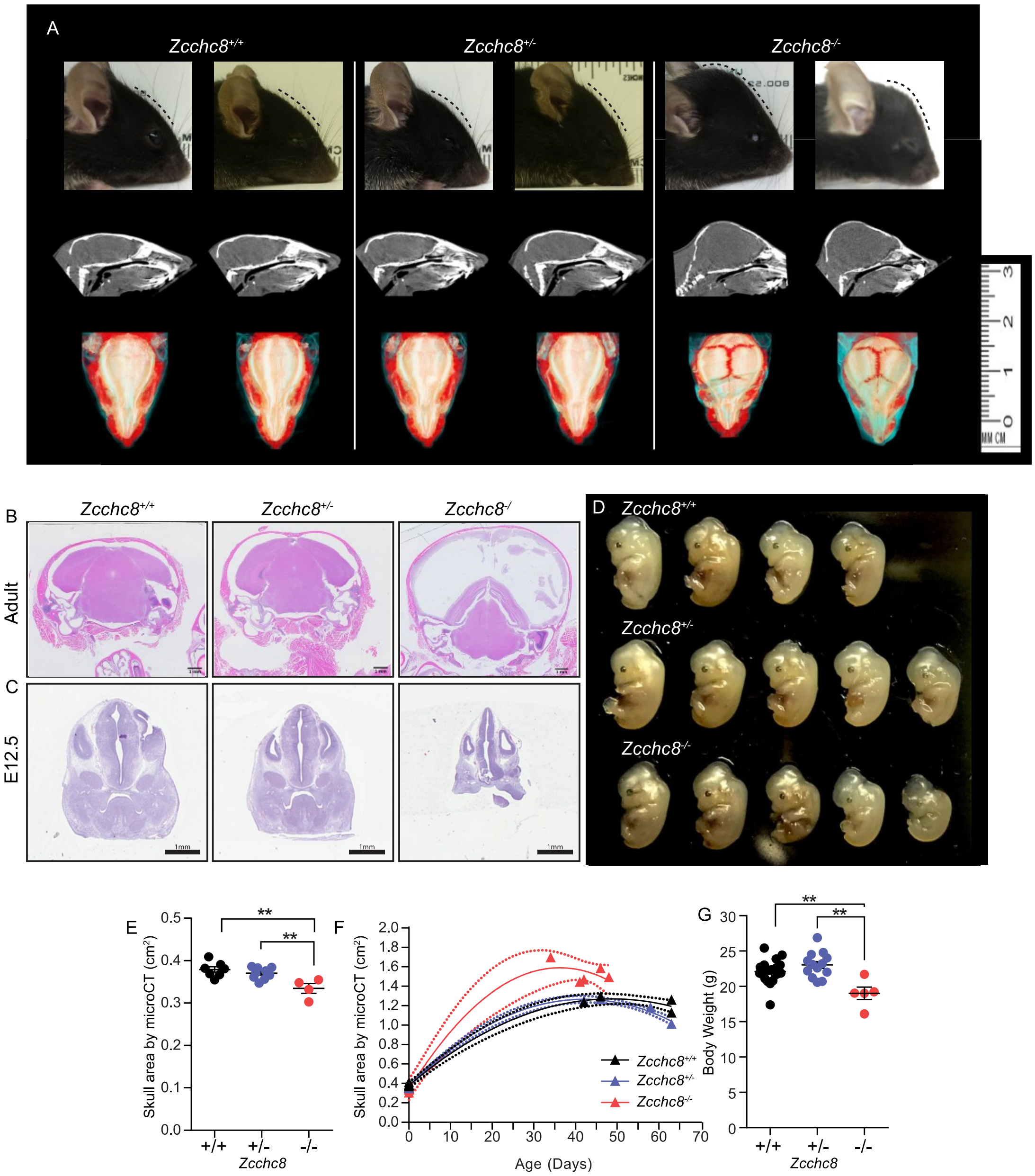


Figure 14. ZCCHC8 loss causes a neurogenesis and hydrocephalus defect in adult knockout mice, but heterozygous mice are intact. A. *(Top row)* Images showing head profile of *Zcchc8* wild-type, heterozygous and homozygous null mice (41-46 days old). The labels show the genotype with a male (left) and a female (right) shown for each genotype. *Zcchc8*$^{-/-}$ mice have abnormal head profile with dome cranium and pointed nose as outlined by the dashed line. *(Middle row)* CT head mid-sagittal images show intact skull shape in *Zcchc8*$^{+/-}$ mice, while *Zcchc8*$^{-/-}$ mice dome-shaped crania. *(Bottom row)* Volume-rendered (VR) CT images of mouse calvaria show widened cranial sutures in *Zcchc8*$^{-/-}$ mice (3 of 6 imaged). None of the *Zcchc8*$^{+/+}$ (0 of 4) or *Zcchc8*$^{+/-}$ mice (0 of 4) imaged had this feature. Each vertical group of images is from the same mouse except the last column (2 different females). B. Representative H&E coronal sections from 8-week

old male heads show no defects in *Zcchc8*$^{+/-}$ (11 examined) compared to *Zcchc8*$^{+/+}$ mice (10 examined). In contrast, *Zcchc8*$^{-/-}$ mice have severe ventricular dilation (seen in 9 of 11 examined). C. E12.5 brain sections show *Zcchc8*$^{-/-}$ have microcephaly but intact brain structures and no ventriculomegaly. D. E12.5 embryos from a single dam show expected Mendelian ratios but *Zcchc8*$^{-/-}$ embryos have small crania. E. Cranial area of P0 pups measured on VR CT images using ImageJ (*Zcch8*$^{+/+}$ n=7, 3M/4F; *Zcch8*$^{+/-}$ n=10, 3M/7F; *Zcch8*$^{-/-}$ n=4, 3M/1F). F. Cranial area as in E with best fit curve (second order polynomial quadratic) showing *Zcchc8*$^{-/-}$ mice become macrocephalic between birth and adulthood reflecting their hydrocephalus (*Zcchc8*$^{+/+}$, n=4, 3M/1F; *Zcchc8*$^{+/-}$, n=4, 3M/1F; *Zcchc8*$^{+/-}$, n=5, 3M/2F). Dashed lines denote the 95% confidence intervals. F. Total body weight of adult male mice (42 to 63 days old, *Zcch8*$^{+/+}$, n=20; *Zcch8*$^{+/-}$, n=13, *Zcch8*$^{-/-}$, n=5). E and F show mean values ±SEM. **P<0.01 (Student's *t*-test, two-sided).

Global RNA analysis shows *Zcchc8*$^{-/-}$ developing brains have defective turnover of RNA polymerase II low abundance transcripts

The mouse phenotypes we documented indicated that, while *Zcchc8* heterozygous loss is well- tolerated, as would be expected for first generation mice that have telomerase insufficiency, complete *Zcchc8* loss caused severe neurodevelopmental defects. To better understand the basis for this in *Zcchc8*$^{-/-}$ adult mice, we performed RNAseq on E12.5 brains. We chose this timepoint in development because the Mendelian ratios were preserved (Figure 13G). We prepared the libraries using an adenylated RNA selection to focus the analysis on protein-coding genes. In a hierarchical analysis, and using no supervision, each of the three *Zcchc8* genotypes clustered separately, but there was an obvious deviation in the *Zcchc8*$^{-/-}$ gene expression pattern from *wild-type* and *heterozygous* brains (Figure 15A). In a volcano plot analysis, we found little difference when we compared *Zcchc8*$^{+/-}$ and *Zcchc8*$^{+/+}$ transcriptomes (Figure 15B). However, the transcriptome from *Zcchc8*$^{-/-}$ embryos showed a skewed and upregulated pattern relative to *wild-types* (Figure 15C). Among these upregulated genes, adenylated *TR* had a linear 6.8-fold higher level relative in *Zcchc8*$^{-/-}$ compared to *wild-type* levels (+5.9 standard deviations in the ordered gene comparisons, $P=1.72\times10^{-8}$, with a one-way two-tailed

ANOVA) despite total TR being down (Figure 13D. We analyzed 3'ends of the *TR* queried in the RNA-seq library and found *Zcchc8*$^{-/-}$ embryos had more extended *TR* forms than *wild-types* and *heterozygotes* (P=0.02 relative to the latter comparison) with *Zcchc8*$^{+/-}$ embryos showing an intermediate phenotype (Figure 15D). Notably, the increase in 3'elongated nascent transcripts was specific to the upregulated genes and was not seen for other RNA pol II genes which had no change in overall levels (e.g. *Arf1* and *Arf3*) nor for RNA pol III non-coding RNAs such as *Rmrp* (Figure 15D). These data suggested that the increased adenylated *TR* levels are in *Zcchc8*$^{-/-}$ are the aberrant forms similar to what we saw in human mutant cells.

We next examined the identity of these most highly upregulated genes in the *Zcchc8*$^{-/-}$ vs. *Zcchc8*$^{+/+}$ analysis (defined as group with >2SD-fold change). *TR* is a low abundance RNA, and we found that the most upregulated genes fell within the lowest FPKM values in the *wild-type* distribution ($P<1x10^{-8}$, Mann-Whitney test and Bootstrap analysis, Figure 15E). This pattern was in contrast to the more broad distribution of the most *downregulated* gene (defined as >2 SD-fold change) in the same *Zcchc8*$^{-/-}$ vs. *Zcchc8*$^{+/+}$ comparison (Figure 15E). These data suggested that ZCCHC8 is involved in the processing and/or turnover of polyadenylated RNAs with relative low abundance in addition to *TR*.

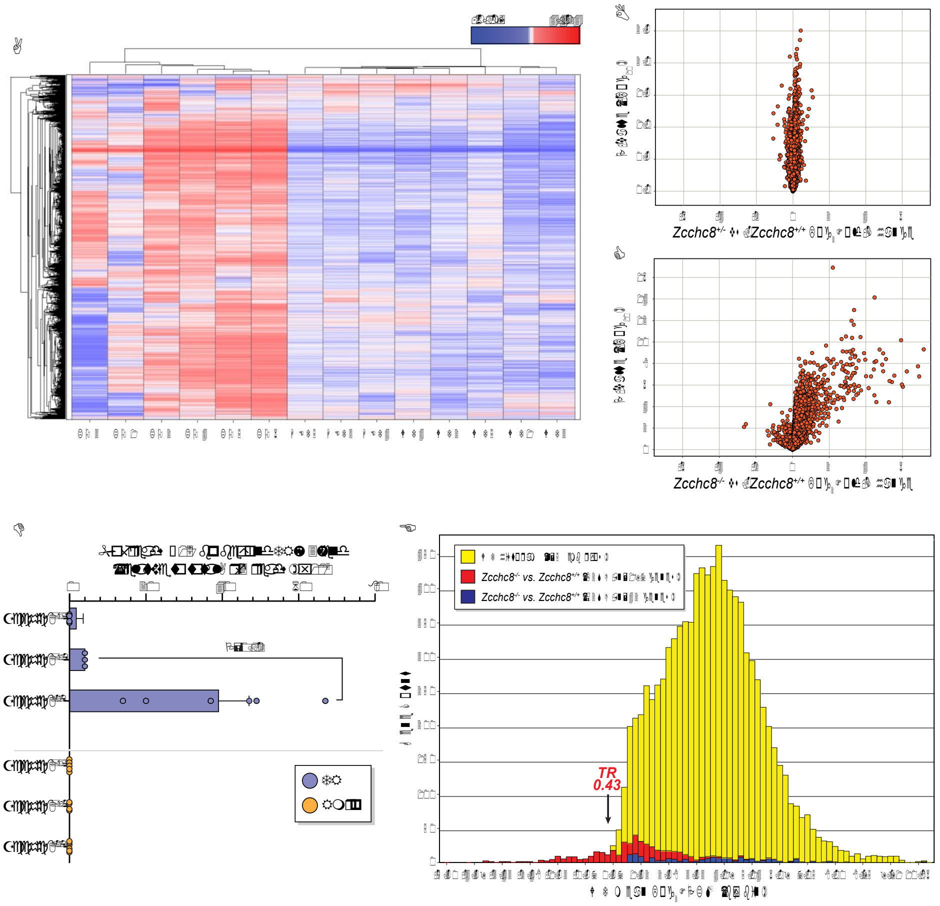
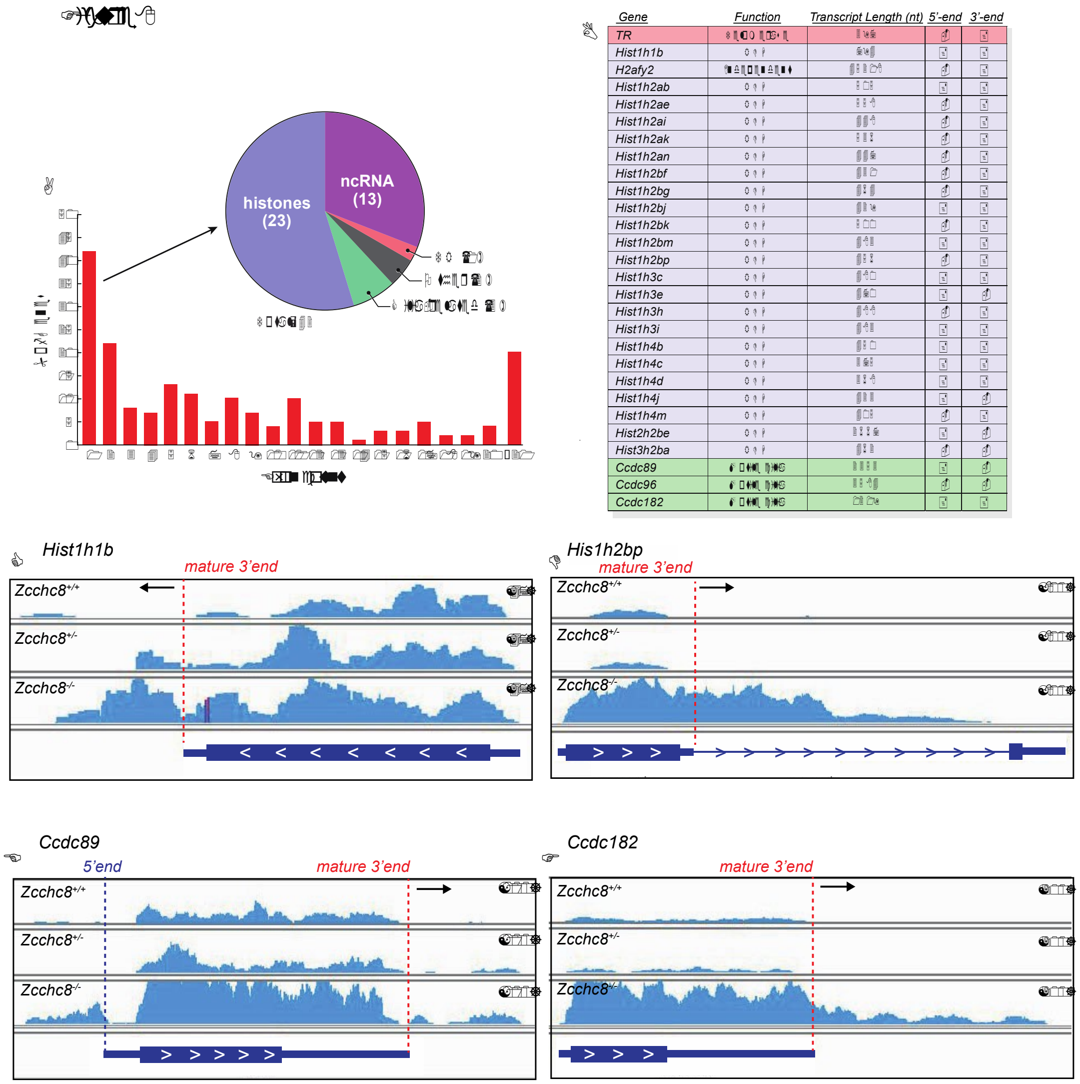


Figure 15. *Zcchc8*$^{-/-}$ mice show an upregulation of low abundance adenylated RNAs including *TR*. A. Heatmap and dendrogram of gene expression showing unsupervised analysis of 9,788 genes. Colors denote mean-subtracted FPKM expression values on a $\log_2$ scale. *Zcchc8*$^{+/+}$ n=5, *Zcchc8*$^{+/-}$ n=3, *Zcchc8*$^{-/-}$, n=6 embryonic brains. Each column is labeled below by WT, HET, KO followed by the embryo number 1,2,3…etc. and refer to respective *Zcchc8* genotypes. The $\log_2$ expression value was subtracted from the mean $\log_2$ expression value of the entire cohort. The dendrogram showing relatedness of the samples is above, and relatedness of the gene transcripts is to the left. The differential change in gene expression is shown as positive and negative change on color scale indicated in key. B and C. Volcano plots depicting the $\log_2$-fold changes (x-axis) vs. $-\log_{10}$ P-values calculated by two-tailed one-way ANOVA of RNA-seq expression data (y-axis) for the *Zcchc8*$^{+/-}$ and *Zcchc8*$^{-/-}$ *vs. Zcchc8*$^{+/+}$ comparisons, respectively. Each dot represents a single transcript. D. Extended reads defined as ≥15 bases beyond the mature RNA end for each of *TR* and *Rmrp* is shown. Data were normalized to total Arf3 reads per embryo and each dot represents values from an embryo. Error bars represent s.e.m and P-value is calculated based on a two-sided Student's t-test. E. Histogram of number of genes at each expression value denoted on the x-axis by the mean $\log_2$FPKM values obtained from *Zcchc8* wild-type embryos (n=5). Genes that have > 2 SD higher levels in the *Zcchc8*$^{+/+}$ vs. *Zcchc8*$^{-/-}$ comparison are shown in red (n=197) and fall on the low abundance end of the histogram with TR and its

mean FPKM in wild-type embryos shown. The downregulated genes defined as <2 SD (n=42) are shown in blue and appear uniformly distributed on the distribution.

Zcchc8 is required for post-transcriptional processing of a subset of replication-dependent histones and some motile cilia components in addition to *TR*

We examined other transcript characteristics of the upregulated, low abundance genes and found that a large subset, 22% (42 of the 188 with known gene structure), were, like *TR*, intronless and transcribed by RNA polymerase II (Figure 16A). Notably, most of these intronless transcripts had similar length to *TR*, between 400 and 500 nucleotides (Figure 16B). When we examined the identity of the intronless genes, we found a striking enrichment of two subsets. The first were histone genes which comprised more than half (23 of 42, 55%). Almost all of these, 22 of 23, were Replication-Dependent Histones (RDH) (Figure 16B). The mouse genome has 65 RDH genes and most of these, 51 of 65 (78%), fall in the major histone cluster on mouse chromosome 13(*89*). Our analysis enriched for nearly half of these (22 of 51 RDH vs. 55 RDH of 9,788 total high quality transcripts, P<0.0001, Fisher's exact test). The second subset involved intronless genes with coiled-coil domains that encode cilia components (Figure 16A-B). Motile cilia on epithelial cells lining the post-natal ventricles in the mouse brain are required for adequate



Gene	Function	Transcript Length (nt)	5'-end	3'-end
TR	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h1b	[illegible]	[illegible]	[illegible]	[illegible]
H2afy2	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2ab	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2ae	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2ai	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2ak	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2an	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2bf	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2bg	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2bj	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2bk	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2bm	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h2bp	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h3c	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h3e	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h3h	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h3i	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h4b	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h4c	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h4d	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h4j	[illegible]	[illegible]	[illegible]	[illegible]
Hist1h4m	[illegible]	[illegible]	[illegible]	[illegible]
Hist2h2be	[illegible]	[illegible]	[illegible]	[illegible]
Hist3h2ba	[illegible]	[illegible]	[illegible]	[illegible]
Ccdc89	[illegible]	[illegible]	[illegible]	[illegible]
Ccdc96	[illegible]	[illegible]	[illegible]	[illegible]
Ccdc182	[illegible]	[illegible]	[illegible]	[illegible]

Figure 16. Telomerase RNA shares a requirement with replication-dependent histones (RDH) and some cilia genes for Zcch8-dependent 3'-end processing. A. Histogram of the most upregulated (>2SD) transcripts in the *Zcchc8*$^{-/-}$ vs. *Zcchc8*$^{+/+}$ by exon number shows the largest category are intronless (42 of 188 with known gene structure, 22%). The pie charts divides the intronless genes by functional category. **B.** Annotation of 28 upregulated intronless genes (*TR*, histones and cilia) shows a majority of the histones represented are RDH (23 of 24, 96%). The vast majority also have an annotated transcript size in the range of *TR* (NCBI Gene database, https://www.ncbi.nlm.nih.gov/gene, accessed 1/1/2019) between 400 and 560 (22 of 28). The 3' and 5'end analyses denote longer ends at each of these mature RNA boundaries in *Zcchc8*$^{-/-}$ as manually visualized in Integrative Genomic Viewer (IGV). Columns refer to 5'end and 3'end refer to

presence of additional reads beyond these boundaries with 5'end reads usually referring to upstream reads that are not necessarily contiguous. **C** and **D.** Genome browser sashimi plots from IGV viewer showing extended 3'ends as labeled above from two representative histone genes *Hist1h1b* and *His1h2bp* in each of *Zcchc8*$^{+/+}$, *Zcchc8*$^{+/-}$, *Zcchc8*$^{-/-}$ transcriptomes, respectively. **E** and **F.** Sashimi plots for two coiled-coil domain containing cilia genes *Ccdc89* and *Ccdc182,* respectively by genotype. For *Ccdc89*, in addition to accumulation of reads beyond the mature 3'-end in the *Zcchc8*$^{-/-}$ transcriptome, there is also an increase in discontinuous upstream of gene 5'end reads that resemble so-called PROMoter uPstream Transcripts (PROMPTs).

cerebrospinal fluid circulation, and their dysfunction could explain the hydrocephalus-otitis phenotype we saw in *Zcchc8*$^{-/-}$ mice. In total, among the most aberrantly processed upregulated polyadenylated genes, there were five coiled-coil domain genes involved in ciliagenesis. In addition, there were transcripts that encode transmembrane proteins (Tmems) that are involved in formation of the ciliary transition zones (n=4). ZCCHC8 was also required for the turnover of *Dnah7b*, an essential component of the dynein axonemal heavy chain of cilia (*90*), as well as the intraflagellar transporter *Ttc26* for which mutations are known to cause hydrocephalus in mice (*91*).

Replication-dependent histones and cilia transcripts show end processing defects like *TR*

We next asked whether the upregulation of RNAs in the *Zcchc8*$^{-/-}$ transcriptome, similar to *TR*, reflects a misprocessing defect. We manually examined the intronless RDH and cilia genes along with *TR* and found that, in addition to their overall increased abundance, most of them (22 of 28) had a 3'end defect marked by genomically-encoded extensions of variable length beyond the annotated 3'end (Figure 16B and examples in Figures 16C-F). This was true for both the RDH and cilia genes we examined (Figures 16B-F). We additionally noted that at least half of these genes (14 of 28) also had accumulated upstream reads that were discontiguous with the 5'gene end and resembled so-called PROMoter uPstream Transcripts (PROMPTS) that have been reported with depletion of NEXT and nuclear RNA exosome components (*92*).

DISCUSSION

This segmental phenotype associated with *TR* insufficiency is evident in the intact phenotype of first generation *Zcchc8*$^{+/-}$ mice relative to *Zcchc8*$^{-/-}$ mice. The transcriptome was an enrichment of dysregulated polyadenylated RNAs and misprocessed genes involved in ciliogenesis and these are potential candidate genes for a secondary ciliopathy in *Zcchc8*$^{-/-}$ mice and suggest a possibility that some RNA dysregulation syndromes, such as those caused by *ZCCHC8* mutations in humans, may overlap in their features with ciliopathies. Our observations are notable since the mouse model we generated may be the first to study the mammalian consequences of mutations in an RNA exosome targeting or an RNA exosome-related gene.

Several RNA dysregulation syndromes have been described and they share defects in neurodevelopment. For example, biallelic mutations in *EXOSC3* have been linked to pontocerebellar hypoplasia, and mutations in *EXOSC8* cause vision and hearing loss, but both of these genes are components of the cytoplasmic as well as nuclear RNA exosomes [reviewed in (*93*)]. Homozygous mutations in *RBM7* have been described in one patient with spinomuscular atrophy who was cognitively intact at 2 years of age (*94*). These phenotypes contrast with the phenotype of autosomal recessive intellectual disability that was found in one of 136 families associated with homozygous null mutations in *ZCCHC8*

(*86*). These observations in light of our data may suggest that neurodevelopment is particularly susceptible to inherited defects in RNA processing genes.

The analysis of RNAs in *Zcchc8*$^{-/-}$ embryos shows that in the homozygous null state, there was dysregulation of low abundance, RNA polymerase II transcripts. Many of these RNAs shared a similar length with *TR*. Low abundance genes were recently identified to be enriched in mouse embryonic stem cells that were depleted for the non-catalytic nuclear RNA exosome component RRP40 (*95*). Our studies in *ZCCHC8* deficient animals, in light of these published data suggest that, through yet-to-be characterized mechanisms, ZCCHC8 and the nuclear RNA exosome preferentially are involved in the processing of low abundance polyadenylated RNAs including *TR*. The RNA species that are preferentially targeted to the RNA exosome by ZCCHC8 have not been previously studied. Prior studies using HeLa cell lines and iCLIP data from over-expression of tagged RBM7, another component of NEXT, found an association with replication dependent histones (RDHs) but suggested a role for the RNA exosome in nascent RDH turnover (*92, 96*). Our data from the developing mouse transcriptome support that but suggest that ZCCHC8 is also required for RDH 3'end processing, in addition to turnover, for at least a subset. Classically, RDH mRNAs are thought to be not polyadenylated but have a unique 3'end stem-loop structure that is required for their endonucleolytic cleavage (*97*). Recent data, however, from transcriptomes of terminally differentiated tissues found that a subset of RDHs are indeed polyadenylated (*98*), similar to our data. Specifically, we find that there are at least some settings where RDH mRNA is adenylated and requires nuclear RNA exosome targeting for its turnover as well as 3'end processing.

Overall, we show that vertebrate *TR* shares a requirement for ZCCHC8 targeting to the nuclear RNA exosome together with at least a subset of RDH genes and other low abundance polyadenylated RNAs. When genetically disturbed, this misprocessing leads to *TR* insufficiency and telomere-mediated disease in heterozygous loss and a more severe neurodevelopmental disease upon total loss of ZCCHC8.

METHODS

Quantitative Real Time-PCR (qRT-PCR) for TR. RNA levels were measured using the SYBR Green qRT-PCR method as previously described (*18*). Briefly, RNA was isolated (RNeasy, Qiagen) and reverse transcribed using random hexamer primers and Superscript III (Invitrogen). Within any given experiment, RNA isolation and cDNA reverse transcription were side-by-side for all the samples. TR levels were normalized to *ARF3* (human) and *Hprt* (mouse). Primer sequences for *mTR* qRT-PCR both total and extended forms available upon request.

Western blots. Total protein was isolated from cells in RIPA buffer (Cell Signaling Technology) with cOmplete Mini protease inhibitor (Sigma-Aldrich) and lysates were quantified using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Using the NuPAGE SDS-PAGE gel system, protein (10-20μg) was run on 10% Bis-Tris gels with MOPS-SDS running buffer at 150V and transferred to a PVDF membrane by wet transfer using XCell SureLock Mini-Cell Elecrophoresis System or by the iBlot 2 Dry Blotting System at 30V (Thermo Fisher Scientific). Immunoblotting was performed using the LI-COR system as previously (*18*). Membranes were blocked for 1h to overnight prior to antibody staining. The following antibodies were used with clone identifiers listed for the two monoclonal antibodies (Actin and Myc): Actin (mouse, ab8226, 1:2000; Abcam, mAbcam8226), Myc (mouse, clone 4A6, 1:1000; Millipore, 4A6), RBM7 (rabbit, HPA013993, 1:250, Sigma), SKIV2L2 (rabbit, ab187884, 1:500, Abcam), Tubulin (rabbit, ab6046, 1:5000; Abcam), ZCCHC8 (mouse, ab68739, 1:500, Abcam). Membranes were stained with anti-mouse or anti-rabbit IRDye secondary antibodies (IR680 or IR800, donkey, 1:10,000 for both) before visualization on an Odyssey scanner (LI-COR). Quantification of protein levels was done using ImageJ (*80*).

Northern quantification of TR. We optimized a protocol for quantifying TR levels using a modified northern protocol to maximize visualization of TR, a low abundance RNA. Total RNA (0.75-6µg) was mixed with RNA loading buffer (Sigma R4268-1VI) and heated to 95°C for 3m then cooled on ice for 3m. Samples were then immediately loaded onto a pre-run Novex 6% TBE-UREA gel and run at 180V in heated TBE buffer (65°C) for 55m on an Isotemp hot plate set to 70°C (Thermo Fisher Scientific). Gels were transferred at 20V onto an Amerhsam Hybond-XL membrane for 90m (GE Healthcare Life Sciences). The cross-linking, hybridization and probe preparation were performed as previously described for *mTR* and *5.8S* (*18*).

Cloning and site directed mutagenesis. Mouse *RBM7* (NM_144948) and mouse *Skiv2l2* (NM_028151) were cloned from total cDNA with addition of an N-terminal Myc-DDK tag into a CMV promoter driven pcDNA5/FRT/TO expression vector using restriction digestion and Gibson cloning (New England Biolabs). Myc-tagged mouse ZCCHC8 cDNA (NM_028151) cDNA was purchased in a pCMV3 expression vector (MG51487-NM, Sino Biological). Plasmids were transfected using Lipofectamine 2000 or 3000 (Invitrogen) and protein lysates were harvested after 24-48h.

Mouse study approval, maintenance and genotyping. The mouse studies were reviewed and approved by the Institutional Animal Care and Use Committee and the procedures conformed with the Guide for the Care and Use of Laboratory Animals (*99*). Mice were housed in the Johns Hopkins University School of Medicine East Baltimore campus. All the mice were on a pure C57BL/6J background. *mTR* null mice were derived and maintained as previously described (*87*). The derivation of *Zcchc8* null mice is described below and these mice were genotyped primers which amplified 363 bp and 396 bp products for the *wild-type* and mutant allele, respectively. Sperm from mice carrying the null *Zcchc8* allele was deposited in the Jackson Laboratories Biorepository, Maine (Stock ID 404814).

Generation of Zcchc8 null mice. We designed a CRISPR/Cas9 strategy to disrupt the *Zcchc8* allele by zygote injection of a sgRNA targeting exon 2 (5' tgaacattctgacaagaccc 3', 20 bp), and a DNA oligo for homology directed repair (HDR) and Cas9 protein. The HDR oligo (5' aagaacttaaaagaaagttgaacattctgacaagaGGATCCGATT ACAAGGACGACGATGACAAGTAGcccaggtatgacatcttgaaattacacacgagctg 3', 103 bp) targeted a 33 bp insertion using 35-base pairing flanking homology arms which introduced a premature stop codon at amino acid 93 (P83delinsGSDYKDDDDK*) as well as a *BamHI* site and FLAG-tag sequence (Figure 5A). To facilitate a genetic screen for the edited alleles, we used DDK sequence specific primers that fell within the FLAG-tag sequence (Figure 5A). Injections into C57BL/6J zygotes were performed at the Johns Hopkins Transgenic Mouse Core Facility. Founder mice were back-crossed to *wildtype* C57BL/6J two times prior to interbreeding of $ZCCHC8^{+/-}$ mice.

Mouse phenotyping and pathology. Mouse ear fibroblasts were derived and maintained as described previously (*18*). Mice were individually euthanized by exposure to gradually increasing concentrations of CO_2 then exsanguination by cardiocentesis and perfusion with heparinized saline, then 10% NBF via left cardiac ventricle. Perfused tissues were fixed in 10% NBF for at least 24h prior to trimming. The head and hindlimb were separately decalcified with Formical 4 (StatLab Medical Products, TX). Gross examination and tissue collection was performed as described in (Brayton 2014 Anaylsis of Phenotype). Sections (~5µm) were stained with hematoxylin and eosin and reviewed by two veterinary pathologists (BK and CB). Image capture was performed on a Nikon 55i microscope and slides were scanned at 20x on an Aperio AT2 instrument (Leica Biosystems, IL).

Computed tomography (CT) imaging and image analysis. Mice were anesthetized using isoflurane prior to and during imaging. Gross images of the mice's heads were taken using a phone camera. CT imaging was performed using the nano PET/CT Small Animal Imager (Mediso) (Johns Hopkins Center for Infection and Inflammation Imaging Research). CT images were visualized on RadiAnt DICOM Viewer (v.4.6.5.18450, 64-bit, Medixant). Using the CT Bone setting, mid-sagittal CT images were acquired using the 3D Multiplanar Reconstruction viewer to align the axes to the middle of the nasal cavity and incisors in the coronal and axial settings respectively. Virtual reconstruction CT images were generated using the 3D Volume Rendering function and saved in the Bones and Skin 3 setting. Cranial area was measured as a surrogate for brain volume using the polygon selection tool of ImageJ. To ensure there are no sex-specific phenotypes, for newborn pup imaging studies, we assessed male/female by PCR as described in (*100*).

RNAseq isolation, library preparation and sequencing. E12.5 brains were isolated and the ftissue was placed into RNAlater (Qiagen) then homogenized in a Bullet Blender using zirconium oxide coated beads (Next Advance). RNA was prepared using RNAeasy kit (Qiagen). Libraries were prepared for RNAseq using the standard protocol for Truseq Stranded mRNA Library preparation (Illumina). The sequencing was completed using NovaSeq S1 flowcell paired end 150 bp. This generated 253 million 150 bp long paired-end reads per sample (range 195-321 million), of which 95.8% on average (range 90.9 to 96.6%) mapped to the mouse genome following the procedure described below.

RNAseq bioinformatics. Transcriptomic data collected by RNA-seq were analyzed to determine the genes that are present in each sample, their expression levels, and the differences between expression levels among different genotypes. *Alignment and annotation.* Following quality checking with the software FastQC (v. 0.10.0)(*101*), reads were end-trimmed to 100 bp and mapped to the mouse genome version mm10 with the alignment tool Tophat2 v.2.1.0 (*102*). Between 91%-97% of reads per sample mapped to the mouse genome, of which a very small fraction (1.5%-1.9%) had multiple matches, and the concordant read fraction was 87.8%-94.4%. The aligned reads were assembled with CLASS2 v.2.1.7 (*103*) to create partial gene and transcript models (transfrags). Transfrags from all samples were further merged with Cuffmerge (v. 2.2.1) (*104*) and mapped to the GENCODE v.M17 (https://www.genecodegenes.org) gene models, to create a unified set of gene annotations for differential analyses. Gene and transcript expression levels (in FPKM) were then computed with the tool Cuffdiff2 v.2.2.1 (*105*) and differentially expressed genes (transcripts) were determined by statistical analysis. The alignment and annotation were performed in August 2018.

Gene expression analysis. Differential gene expression analyses were performed using the FPKM files comprising 67,811 transcripts each generated as described above for each embryonic sample. The transcript identifiers were derived from these FPKM files' "gene" column and updated to current nomenclature. The raw FPKM values of 0.0 were treated as nulls and actual values were transformed into log2 annotation and quantile normalized across the 14 samples. The three biological classes', KO, HET, and WT, normalized signals underwent differential expression analysis with a two-tailed one-way *t*-test ANOVA using the Partek GS 6.16.0812 platform. A standard deviation analysis was performed for each class-class comparison using those transcripts that had high quality data: an NCBI Entrez gene ID, a mean FPKM linear value >2.0 in at least one cell class, and a non-zero FPKM signal value from all the 14 samples. On the order of 9.5K transcripts met these three criteria for each cell-class comparison and were used for subsequent downstream functional analyses. The raw data were deposited in the National Center for Biotechnology Information Gene Expression Omnibus database GEO (#GSE126108).

RNAseq 3'-end analysis. To elucidate the presence of 3' extensions for *mTR* and *mRmrp*, and for relative abundance comparisons among *Zcchc8*$^{+/+}$, *Zcchc8*$^{+/-}$, and *Zcchc8*$^{-/-}$ mice (14 samples), we extracted RNA-seq read pairs partially aligning to the gene from the mouse RNA-seq Fastq files generated as part of the RNAseq bioinformatic analysis. Reads were then processed using the protocol described for 3'RACE-seq data with some modifications, described herein. Briefly, reads (and fragments) mapping to the target gene were determined with Bowtie2 (*82*) with the option *'-local'*. Reads were then trimmed of Illumina adapters using the tool cutadapt (*106*), and an artificial 'tag' consisting of the 28 bp linker sequence ACTGTAGGCACCATCAATCGTTACGTAG was added to the 3' ends of forward mapping reads, whereas the reverse complement linker sequence was pre-pended to the 5' end of the reverse mapping reads. Similarly, the linker sequence was appended to the 3'end of the gene(s) to generate reference sequences. Unlike with 3'RACE-seq reads, where the start of the linker marked the precise position of the gene's end, here the linker merely indicates that the end of the gene is at or downstream of the mapping location of the read. Hence, reads spanning the gene-linker junction collectively encompass 3' RNA extensions, both adenylated and genomic, as well as potential genomic and other types of contamination. The relative abundance of these reads was then compared across the genotypes.

Statistical analyses. Singlepoint parametric linkage was analyzed using logarithm of the odds scores. qRT-PCR analyses were reported as standard error of the mean (s.e.m.), and *P* values were calculated using GraphPad Prism software. For each PCR run, samples were done in triplicate. All P-values shown are two-sided. Statistical tests uses are indicated in the text and legends with the respective data.

References

1. C. W. Greider, E. H. Blackburn, Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405-413 (1985).
2. C. W. Greider, E. H. Blackburn, A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**, 331-337 (1989).
3. G. Cristofari, J. Lingner, Telomere length homeostasis requires that telomerase levels are limiting. *Embo J* **25**, 565-574 (2006).
4. C. W. Greider, Telomerase RNA levels limit the telomere length equilibrium. *Cold Spring Harb Symp Quant Biol* **71**, 225-229 (2006).
5. N. Erdmann, Y. Liu, L. Harrington, Distinct dosage requirements for the maintenance of long and short telomeres in mTert heterozygous mice. *Proc Natl Acad Sci U S A* **101**, 6080-6085 (2004).
6. L. Y. Hao *et al.*, Short telomeres, even in the presence of telomerase, limit tissue renewal capacity. *Cell* **123**, 1121-1131 (2005).
7. A. D. Mozdy, T. R. Cech, Low abundance of telomerase in yeast: implications for telomerase haploinsufficiency. *Rna* **12**, 1721-1737 (2006).
8. T. Vulliamy *et al.*, The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* **413**, 432-435 (2001).
9. M. Y. Armanios *et al.*, Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* **356**, 1317-1326 (2007).
10. M. A. Strong *et al.*, Phenotypes in mTERT(+)/(-) and mTERT(-)/(-) mice are due to short telomeres, not telomere-independent functions of telomerase reverse transcriptase. *Mol Cell Biol* **31**, 2369-2379 (2011).
11. M. Armanios, Telomerase and idiopathic pulmonary fibrosis. *Mutation research* **730**, 52-58 (2012).
12. L. Ungar *et al.*, A genome-wide screen for essential yeast genes that affect telomere length maintenance. *Nucleic Acids Res* **37**, 3840-3849 (2009).
13. S. H. Askree *et al.*, A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. *Proc Natl Acad Sci U S A* **101**, 8658-8663 (2004).
14. T. Gatbonton *et al.*, Telomere length as a quantitative trait: genome-wide survey and genetic mapping of telomere length-control genes in yeast. *PLoS Genet* **2**, e35 (2006).
15. K. S. Hathcock *et al.*, Haploinsufficiency of mTR results in defects in telomere elongation. *Proc Natl Acad Sci U S A* **99**, 3591-3596 (2002).
16. B. D. Stuart *et al.*, Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat Genet* **47**, 512-517 (2015).
17. D. H. Moon *et al.*, Poly(A)-specific ribonuclease (PARN) mediates 3'-end maturation of the telomerase RNA component. *Nat Genet* **47**, 1482-1488 (2015).
18. S. E. Stanley *et al.*, Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. *Science translational medicine* **8**, 351ra107 (2016).
19. H. Tummala *et al.*, Poly(A)-specific ribonuclease deficiency impacts telomere biology and causes dyskeratosis congenita. *J Clin Invest* **125**, 2151-2160 (2015).
20. C. W. Greider, E. H. Blackburn, The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887-898 (1987).
21. U. T. Meier, The many facets of H/ACA ribonucleoproteins. *Chromosoma* **114**, 1-14 (2005).

22. A. Virtanen, N. Henriksson, P. Nilsson, M. Nissbeck, Poly(A)-specific ribonuclease (PARN): an allosterically regulated, processive and mRNA cap-interacting deadenylase. *Crit Rev Biochem Mol Biol* **48**, 192-209 (2013).
23. J. A. Box, J. T. Bunch, W. Tang, P. Baumann, Spliceosomal cleavage generates the 3' end of telomerase RNA. *Nature* **456**, 910-914 (2008).
24. N. Jamonnak *et al.*, Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* **17**, 2011-2025 (2011).
25. S. Coy, A. Volanakis, S. Shah, L. Vasiljeva, The Sm complex is required for the processing of non-coding RNAs by the exosome. *PLoS One* **8**, e65606 (2013).
26. Y. Vasianovich, R. J. Wellinger, Life and Death of Yeast Telomerase RNA. *Journal of molecular biology* **429**, 3242-3254 (2017).
27. E. D. Egan, K. Collins, Biogenesis of telomerase ribonucleoproteins. *Rna* **18**, 1747-1759 (2012).
28. H. Berndt *et al.*, Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* **18**, 958-972 (2012).
29. K. C. Goldfarb, T. R. Cech, 3' terminal diversity of MRP RNA and other human noncoding RNAs revealed by deep sequencing. *BMC molecular biology* **14**, 23 (2013).
30. D. Nguyen *et al.*, A Polyadenylation-Dependent 3' End Maturation Pathway Is Required for the Synthesis of the Human Telomerase RNA. *Cell Rep* **13**, 2244-2257 (2015).
31. J. Feng *et al.*, The RNA component of human telomerase. *Science* **269**, 1236-1241 (1995).
32. C. S. Hinkley *et al.*, The mouse telomerase RNA 5"-end lies just upstream of the telomerase template sequence. *Nucleic Acids Res* **26**, 532-536 (1998).
33. D. Fu, K. Collins, Human telomerase and Cajal body ribonucleoproteins share a unique specificity of Sm protein association. *Genes Dev* **20**, 531-536 (2006).
34. C. Girard *et al.*, Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *J Biol Chem* **283**, 2060-2069 (2008).
35. C. Verheggen *et al.*, Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J* **21**, 2736-2745 (2002).
36. S. Giacometti *et al.*, Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate. *Cell Rep* **18**, 2635-2650 (2017).
37. M. Lubas *et al.*, Interaction profiling identifies the human nuclear exosome targeting complex. *Molecular cell* **43**, 624-637 (2011).
38. K. Winczura *et al.*, Characterizing ZC3H18, a Multi-domain Protein at the Interface of RNA Production and Destruction Decisions. *Cell Rep* **22**, 44-58 (2018).
39. S. Boulon *et al.*, PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Molecular cell* **16**, 777-787 (2004).
40. C. K. Tseng *et al.*, Human Telomerase RNA Processing and Quality Control. *Cell Rep* **13**, 2232-2243 (2015).
41. S. E. Stanley, M. Armanios, The short and long telomere syndromes: paired paradigms for molecular medicine. *Curr Opin Genet Dev* **33**, 1-9 (2015).
42. S. E. Stanley *et al.*, Telomerase mutations in smokers with severe emphysema. *J Clin Invest* **125**, 563-570 (2015).
43. X. Darzacq *et al.*, Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *The Journal of cell biology* **173**, 207-218 (2006).
44. C. Hoareau-Aveilla, M. Bonoli, M. Caizergues-Ferrer, Y. Henry, hNaf1 is required for accumulation of human box H/ACA snoRNPs, scaRNPs, and telomerase. *RNA* **12**, 832-840 (2006).
45. Y. T. Yu, U. T. Meier, RNA-guided isomerization of uridine to pseudouridine--pseudouridylation. *RNA Biol* **11**, 1483-1494 (2014).

46. T. M. Carlile *et al.*, Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143-146 (2014).
47. J. R. Mitchell, J. Cheng, K. Collins, A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol Cell Biol* **19**, 567-576 (1999).
48. A. A. Lukowiak, A. Narayanan, Z. H. Li, R. M. Terns, M. P. Terns, The snoRNA domain of vertebrate telomerase RNA functions to localize the RNA within the nucleus. *RNA* **7**, 1833-1844 (2001).
49. J. K. Alder *et al.*, Ancestral mutation in telomerase causes defects in repeat addition processivity and manifests as familial pulmonary fibrosis. *PLoS genetics* **7**, e1001352 (2011).
50. T. J. Vulliamy *et al.*, Differences in disease severity but similar telomere lengths in genetic subgroups of patients with telomerase and shelterin mutations. *PloS one* **6**, e24383 (2011).
51. I. Dokal, T. Vulliamy, Dyskeratosis congenita: its link to telomerase and aplastic anaemia. *Blood Rev* **17**, 217-225 (2003).
52. H. Ly *et al.*, Functional characterization of telomerase RNA variants found in patients with hematologic disorders. *Blood* **105**, 2332-2339 (2005).
53. H. Yamaguchi *et al.*, Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome. *Blood* **102**, 916-918 (2003).
54. K. R. Prowse, C. W. Greider, Developmental and tissue-specific regulation of mouse telomerase and telomere length. *Proc Natl Acad Sci U S A* **92**, 4818-4822 (1995).
55. M. Armanios *et al.*, Short telomeres are sufficient to cause the degenerative defects associated with aging. *Am J Hum Genet* **85**, 823-832 (2009).
56. Y. Mochizuki, J. He, S. Kulkarni, M. Bessler, P. J. Mason, Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *Proc Natl Acad Sci U S A* **101**, 10756-10761 (2004).
57. B. Gu, M. Bessler, P. J. Mason, Dyskerin, telomerase and the DNA damage response. *Cell Cycle* **8**, 6-10 (2009).
58. J. He *et al.*, Targeted disruption of Dkc1, the gene mutated in X-linked dyskeratosis congenita, causes embryonic lethality in mice. *Oncogene* **21**, 7740-7744 (2002).
59. D. Ruggero *et al.*, Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science* **299**, 259-262 (2003).
60. J. D. Podlevsky, C. J. Bley, R. V. Omana, X. Qi, J. J. Chen, The telomerase database. *Nucleic Acids Res* **36**, D339-343 (2008).
61. J. Lingner *et al.*, Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561-567 (1997).
62. M. Armanios, E. H. Blackburn, The telomere syndromes. *Nature reviews. Genetics* **13**, 693-704 (2012).
63. E. M. Parry, J. K. Alder, X. Qi, J. J. Chen, M. Armanios, Syndrome complex of bone marrow failure and pulmonary fibrosis predicts germline defects in telomerase. *Blood* **117**, 5607-5611 (2011).
64. K. D. Tsakiri *et al.*, Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci U S A* **104**, 7552-7557 (2007).
65. J. K. Alder *et al.*, Telomere Phenotypes in Females with Heterozygous Mutations in the Dyskeratosis Congenita 1 (DKC1) Gene. *Hum Mutat*, (2013).
66. J. K. Alder *et al.*, Telomere dysfunction causes alveolar stem cell failure. *Proc Natl Acad Sci U S A* **112**, 5099-5104 (2015).
67. J. D. Cogan *et al.*, Rare Variants in RTEL1 are Associated with Familial Interstitial Pneumonia. *Am J Respir Crit Care Med*, (2015).
68. J. R. Mitchell, E. Wood, K. Collins, A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402**, 551-555 (1999).

69. J. L. Chen, M. A. Blasco, C. W. Greider, Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503-514 (2000).
70. T. Vulliamy *et al.*, Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. *Nat Genet*, (2004).
71. J. D. Podlevsky, J. J. Chen, Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol* **13**, 720-732 (2016).
72. M. Armanios *et al.*, Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proc Natl Acad Sci U S A* **102**, 15960-15964 (2005).
73. B. Glaunsinger, D. Ganem, Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover. *Molecular cell* **13**, 713-723 (2004).
74. J. K. Alder *et al.*, Diagnostic utility of telomere length testing in a hospital-based setting. *Proc Natl Acad Sci U S A* **115**, E2358-E2365 (2018).
75. M. B. Penno, M. Pedrotti-Krueger, T. Ray, Cyropreservation of whole blood and isolated lymphocytes for B-cell immortalization. *J. Tiss. Cult. Meth.* **15**, 43-48 (1993).
76. G. M. Baerlocher, I. Vulto, G. de Jong, P. M. Lansdorp, Flow cytometry and FISH to measure the average length of telomeres (flow FISH). *Nat Protoc* **1**, 2365-2376 (2006).
77. A. Hamosh *et al.*, PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum Mutat* **34**, 566-571 (2013).
78. G. R. Abecasis, S. S. Cherny, W. O. Cookson, L. R. Cardon, Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97-101 (2002).
79. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).
80. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).
81. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308 (2013).
82. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
83. B. Walenz, L. Florea, Sim4db and Leaff: utilities for fast batch spliced alignment and sequence indexing. *Bioinformatics* **27**, 1869-1870 (2011).
84. N. W. Kim *et al.*, Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-2015 (1994).
85. M. T. Hemann, C. W. Greider, Wild-derived inbred mouse strains have short telomeres. *Nucleic Acids Res* **28**, 4474-4478 (2000).
86. H. Najmabadi *et al.*, Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57-63 (2011).
87. M. A. Blasco *et al.*, Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25-34 (1997).
88. M. Kousi, N. Katsanis, The Genetic Basis of Hydrocephalus. *Annu Rev Neurosci* **39**, 409-435 (2016).
89. W. F. Marzluff, P. Gongidi, K. R. Woods, J. Jin, L. J. Maltais, The human and mouse replication-dependent histone genes. *Genomics* **80**, 487-498 (2002).
90. Y. J. Zhang *et al.*, Identification of dynein heavy chain 7 as an inner arm component of human cilia that is synthesized but not assembled in a case of primary ciliary dyskinesia. *J Biol Chem* **277**, 17906-17915 (2002).
91. R. E. Swiderski, Y. Nakano, R. F. Mullins, S. Seo, B. Banfi, A mutation in the mouse ttc26 gene leads to impaired hedgehog signaling. *PLoS Genet* **10**, e1004689 (2014).
92. P. R. Andersen *et al.*, The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nature structural & molecular biology* **20**, 1367-1376 (2013).

93. D. J. Morton *et al.*, The RNA exosome and RNA exosome-linked disease. *RNA* **24**, 127-142 (2018).
94. M. Giunta *et al.*, Altered RNA metabolism due to a homozygous RBM7 mutation in a patient with spinal motor neuropathy. *Hum Mol Genet* **25**, 2985-2996 (2016).
95. M. Lloret-Llinares *et al.*, The RNA exosome contributes to gene expression regulation during stem cell differentiation. *Nucleic Acids Res* **46**, 11502-11513 (2018).
96. M. Lubas *et al.*, The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* **10**, 178-192 (2015).
97. W. F. Marzluff, E. J. Wagner, R. J. Duronio, Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature reviews. Genetics* **9**, 843-854 (2008).
98. S. M. Lyons *et al.*, A subset of replication-dependent histone mRNAs are expressed as polyadenylated RNAs in terminally differentiated tissues. *Nucleic Acids Res* **44**, 9190-9205 (2016).
99. in *Guide for the Care and Use of Laboratory Animals,* th, Ed. (Washington (DC), 2011).
100. L. McFarlane, V. Truong, J. S. Palmer, D. Wilhelm, Novel PCR assay for determining the genetic sex of mice. *Sex Dev* **7**, 207-211 (2013).
101. Fastqc package for quality checking of reads. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
102. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
103. L. Song, S. Sabunciyan, L. Florea, CLASS2: accurate and efficient splice variant annotation from RNA-seq reads. *Nucleic Acids Res* **44**, e98 (2016).
104. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-515 (2010).
105. C. Trapnell *et al.*, Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* **31**, 46-53 (2013).
106. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet.journal* **17**, 10-12 (2011).

Curriculum Vitae

Dustin L. Gable
Born in Miller City, Ohio

EDUCATION

2011-present	Medical Scientist Training Program Human Genetics Graduate Program Johns Hopkins University School of Medicine, Baltimore, MD
2007-2011	Bachelor of Science in Allied Health Professions Biomedical Sciences, *summa cum laude* The Ohio State University, Columbus, OH

RESEARCH

2013-present	Human Genetics Graduate Program Department of Oncology & McKusick-Nathans Institute of Genetic Medicine Johns Hopkins University School of Medicine, Baltimore, MD Advisor: Mary Armanios, MD Thesis: Mechanisms of telomerase RNA insufficiency in disease
2007-2011 Genetics	Department of Molecular Virology, Immunology, and Medical The Ohio State University, Columbus, OH Advisor: Amanda Toland, PhD Honors Thesis: The role of microRNA-1 in cutaneous squamous cell carcinoma
June-Aug 2010	Pediatric Oncology Education Program Advisor: Joseph Opferman, PhD Project: MCL-1 in B-Lineage acute lymphoblastic leukemia St. Jude Children's Research Hospital, Memphis, TN

PUBLICATIONS

1. **Gable DL**, Gaysinskaya V, Atik CC, Conover Talbot, C Jr., Kang B, Stanley SE, Pugh, EW, Amat-Codina N, Mirski/Schenk K, Arcasoy MO, Brayton C, Florea L, Armanios M. "A Nuclear RNA exosome targeting component is mutated in familial pulmonary fibrosis and is required for telomerase RNA maturation," *In submission.* 2019 April.
I identified the disease gene and generated the null cell lines and mice and performed other experiments

2. Stanley SE †, **Gable DL†,** Wagner CL, Carlile TM, Hanumanthu VS, Podlevsky JD, Khalil SE, DeZern AE, Rojas-Duran MF, Applegate CD, Alder JK, Parry EM, Gilbert WV, Armanios M. “Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema," *Science Translational Medicine.* 2016 Aug 10. 8(351):351ra107.
†Shared first author.
• I used CRISPR/Cas9 editing to generate and characterize Naf1 knockout mice and cell lines with patient NAF1 mutations.

3. Parry EM, **Gable DL**, Stanley SE, Khalil SE, Antonescu V, Florea L, Armanios M. “Germline Mutations in DNA repair genes in lung adenocarcinoma,” *Journal of Thoracic Oncology.* 2017 Nov;12(11):1673-1678
• I downloaded and supported in all the analyses of the data for this paper and also created summary figures for the paper.

4. Stanley SE, Rao AD, **Gable DL**, McGrath-Morrow S, Armanios M. “Radiation Sensitivity and Radiation Necrosis in the Short Telomere Syndromes,” *Int J Radiat Oncol Biol Phys*. 2015 December 1. 93(5): 1115-1117.
• I brought and validated the radiation cell survival assay used in this paper.

5. Wagner CL, Hanumanthu VS, Talbot CC Jr, Abraham RS, Hamm D, **Gable DL**, Kanakry CG, Applegate CD, Siliciano J, Jackson JB, Desiderio S, Alder JK, Luznik L, Armanios M. “Short telomere syndromes cause a primary T cell immunodeficiency.” *Journal of Clinical Investigation.* 2018 Dec 3;128(12):5222-5234.
• I performed and validated T cell proliferation and apoptosis assays for this paper.

6. You J, Sobreira NL, **Gable DL**, Jurgens J, Grange DK, Belnap N, Siniard A, Szelinger S, Schrauwen I, Richholt RF, Vallee SE, Dinulos MBP, Valle D, Armanios M, Hoover-Fong J. “A Syndromic Intellectual Disability Disorder Caused by Variants in TELO2, a Gene Encoding a Component of the TTT Complex,” *The American Journal of Human Genetics.* 2016 May 5;98(5):909-918
• I performed the Western blots to measure the levels of other PI3 kinases in TELO-2 mutant patient-derived cell lines.

7. Fleming JL†, **Gable DL**†, Samadzadeh-Tarighat S, Cheng L, Yu L, Gillespie JL, Toland AE. “Differential expression of miR-1, a putative tumor suppressing microRNA, in cancer resistant and cancer susceptible mice,” *PeerJ.* 2013 Apr 16;1:e68.
†Shared first author.
• I conducted cell survival and apoptosis studies and identified microRNA target proteins in miR-1 knockdown cell lines.

PRESENTATIONS

Gable DL, Stanley SE, Carlile TM, Hanumanthu VS, Alder JK, Gilbert WV, Armanios M. “Haploinsufficiency of the H/ACA biogenesis factor *NAF1* selectively disrupts telomere length maintenance while sparing ribosomal RNA functions” *Cold Spring Harbor Telomeres and Telomerase Meeting.* 2017 May 4. Abstract and Poster

TEACHING & WORK EXPERIENCE

2003-2010 Gable Dairy Farmhand and Construction Worker
Leipsic, OH (20-70 hours/week)
I was officially employed during middle school through college summers working on my family's dairy farm to harvest crops, milk cows, maintain livestock and building local barns and sheds.

2014-present Post doctoral fellows (2)
Graduate rotation students (3)
Undergraduate students (2)
I assisted new clinical and other post-doctoral fellows and students with their new projects in the lab including experimental design, troubleshooting, and data interpretation.

Jan-May 2016 The Healer's Art Co-instructor
Johns Hopkins School of Medicine
I led activities and discussions for medical students in a curriculum focused on examining clarification, professionalism, and values as a physician.

2003-2010 Gable Dairy Farmhand and Construction
Leipsic, OH (20-70 hours/week)
I was officially employed during middle school through college summers and breaks working on my family's dairy farm to harvest crops, milk cows, maintain livestock and building barns and sheds

LEADERSHIP & VOLUNTEER

2013-2016 Co-President MD/PhD Student Advisory Board
Johns Hopkins Medical Scientist Training Program
• *I was selected by a student-run committee to lead and organize events including retreats, interviews, second look weekend, and medical school-to-research transitions.*

2012-2015 Medical Student Senate Class Representative
Johns Hopkins School of Medicine
• *I was elected as one of four representatives by my medical school class to organize school-wide events, sit on faculty and student committees, and implement change productively.*

2014-2015 Medical School Admissions Committee
Johns Hopkins School of Medicine
• *I was elected as one of sixteen students in the medical school class to conduct weekly interviews for medical school applicants and present them at admissions committee meetings.*

2011-2012 Medical School Second Look Weekend Committee
Oct-April Johns Hopkins School of Medicine
• I was one of six students in medical school class to organize events, speakers, travel, and tours for the revisiting accepted applicants across a four day visit at Hopkins.

2014-2017 Baltimore Community Softball (Team Coach and Captain 2017)
• I played and coached on a community summer softball team and while captain, organized schedules, taught softball skills to players new to the sport, and coached twice weekly games.

2017 Our Daily Bread
Catholic Charities of Baltimore, MD
• Kitchen aid and cook

AWARDS

2009-2010 Goldwater Scholar
• I was one of three from Ohio State selected as a national Scholar ($7500/year) from over 1100 undergraduates nationwide in the sciences, engineering, and mathematics.

2017-2018 Turock Scholar
• I was awarded a research scholarship as the senior student in the Armanios Lab.